

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number  
**WO 01/46258 A2**(51) International Patent Classification<sup>7</sup>: **C07K 14/705**(21) International Application Number: **PCT/US00/35095**(22) International Filing Date:  
22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

60/172,000	23 December 1999 (23.12.1999)	US
60/176,083	14 January 2000 (14.01.2000)	US
60/177,332	21 January 2000 (21.01.2000)	US
60/178,572	28 January 2000 (28.01.2000)	US
60/179,758	2 February 2000 (02.02.2000)	US
60/181,625	10 February 2000 (10.02.2000)	US

## (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/172,000 (CIP)
Filed on	23 December 1999 (23.12.1999)
US	60/176,083 (CIP)
Filed on	14 January 2000 (14.01.2000)
US	60/177,332 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	60/178,572 (CIP)
Filed on	28 January 2000 (28.01.2000)
US	60/179,758 (CIP)
Filed on	2 February 2000 (02.02.2000)
US	60/181,625 (CIP)
Filed on	10 February 2000 (10.02.2000)

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005

(US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **NGUYEN, Dannel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **KHAN, Farrah, A.** [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

## Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **TRANSPORTERS AND ION CHANNELS**

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 01/46258 A2

## TRANSPORTERS AND ION CHANNELS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, and immunological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

### BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including  $K^+$ ,  $NH_4^+$ ,  $P_i$ ,  $SO_4^{2-}$ , sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous  $Na^+/K^+$  ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging techniques and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc.

Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H<sup>+</sup>-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H<sup>+</sup>-linked monocarboxylate transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K<sub>m</sub> values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na<sup>+</sup>-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective transporters for organic cations and organic anions in organs including the kidney, intestine and liver. Organic anion transporters are selective for

hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) *Am. J. Physiol.* 264:C761-C782; Price, N.T. et al. (1998) *Biochem. J.* 329:321-328; and Martinelle, K. and I. Haggstrom (1993) *J. Biotechnol.* 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) *Meth. Enzymol.* 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) *J. Med. Genet.* 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process. Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments,



is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

5 Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a  
10 protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) \*275000 Graves Disease).

15 This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int.  
20 Med. 245:637-642).

### **Ion Channels**

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and  
25 gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration  
30 gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

### **Ion Transporters**

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient. These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion

transporters, including Na<sup>+</sup>-K<sup>+</sup> ATPase, Ca<sup>2+</sup>-ATPase, and H<sup>+</sup>-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na<sup>+</sup> and Ca<sup>2+</sup> are low and cytosolic concentration of K<sup>+</sup> is high. The vacuolar (V) class of ion transporters includes H<sup>+</sup> pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H<sup>+</sup> pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P<sub>i</sub>).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the V<sub>1</sub> domain, a peripheral complex responsible for ATP hydrolysis; and the V<sub>0</sub> domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F<sub>0</sub> domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V<sub>0</sub> domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgacs, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na<sup>+</sup> down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca<sup>2+</sup> out of the cell with transport of Na<sup>+</sup> into the cell (antiport).

#### Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup>, and Cl<sup>-</sup> channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated

chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

5 Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both  $\text{Ca}^{2+}$  and  $\text{Na}^{+}$  (Suzuki, M. et al. (1999) J. Biol. Chem. 10 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the  $\text{Na}^{+}$  and  $\text{Ca}^{2+}$  subfamilies, this domain is repeated four times, while in the  $\text{K}^{+}$  15 channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of  $\text{K}^{+}$  channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated  $\text{Na}^{+}$  and  $\text{K}^{+}$  channels are necessary for the function of electrically excitable cells, 20 such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to  $\text{Na}^{+}$  and  $\text{K}^{+}$  ions. Depolarization of the membrane beyond the threshold level opens voltage-gated  $\text{Na}^{+}$  channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated  $\text{Na}^{+}$  channels, which propagates the depolarization down the length of the cell. Depolarization also opens 25 voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug 30 that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated  $\text{Na}^{+}$  channels are heterotrimeric complexes composed of a 260 kDa pore-forming  $\alpha$  subunit that associates with two smaller auxiliary subunits,  $\beta 1$  and  $\beta 2$ . The  $\beta 2$  subunit is a integral membrane glycoprotein that contains an extracellular Ig domain, and its association with  $\alpha$  and  $\beta 1$

subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) *Cell* 83:433-442).

Non voltage-gated  $\text{Na}^+$  channels include the members of the amiloride-sensitive  $\text{Na}^+$  channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial  $\text{Na}^+$  channel (ENaC) involved in  $\text{Na}^+$  reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized  $\text{H}^+$ -gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric  $\text{Na}^+$ -permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from *C. elegans*. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) *Curr. Opin. Neurobiol.* 8:418-424; Eglén, R.M. et al. (1999) *Trends Pharmacol. Sci.* 20:337-342).

$\text{K}^+$  channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as  $\text{Ca}^{2+}$  and cAMP. In non-excitabile tissue,  $\text{K}^+$  channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes,  $\text{K}^+$  channels are responsible for setting resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a  $\text{Na}^+$ - $\text{K}^+$  pump and ion channels that provide the redistribution of  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Cl}^-$ . The pump actively transports  $\text{Na}^+$  out of the cell and  $\text{K}^+$  into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow  $\text{K}^+$  and  $\text{Cl}^-$  to flow by passive diffusion. Because of the high negative charge within the cytosol,  $\text{Cl}^-$  flows out of the cell. The flow of  $\text{K}^+$  is balanced by an electromotive force pulling  $\text{K}^+$  into the cell, and a  $\text{K}^+$  concentration gradient pushing  $\text{K}^+$  out of the cell. Thus, the resting membrane potential is primarily regulated by  $\text{K}^+$  flow (Salkoff, L. and T. Jegla (1995) *Neuron* 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic  $\beta$  subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated  $\text{K}^+$  channels as well as the delayed rectifier type channels such as the human ether-a-go-go

related gene (HERG) associated with long QT, a cardiac dysrhythmia syndrome (Curran, M.E. (1998) *Curr. Opin. Biotechnol.* 9:565-572; Kaczarowski, G.J. and M.L. Garcia (1999) *Curr. Opin. Chem. Biol.* 3:448-458).

A second superfamily of K<sup>+</sup> channels is composed of the inward rectifying channels (Kir).

5 Kir channels have the property of preferentially conducting K<sup>+</sup> currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K<sup>+</sup> channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker  
10 activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) *Curr. Opin. Neurobiol.* 5:268-277; Curran, *supra*).

The recently recognized TWIK K<sup>+</sup> channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential  
15 in a large set of cell types (Duprat, F. et al. (1997) *EMBO J* 16:5464-5471).

The voltage-gated Ca<sup>2+</sup> channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca<sup>2+</sup> channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type  
20 channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca<sup>2+</sup> channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The  $\alpha_1$  subunit forms the membrane pore and voltage sensor, while the  $\alpha_2\delta$  and  $\beta$  subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel.  
25 These subunits are encoded by at least six  $\alpha_1$ , one  $\alpha_2\delta$ , and four  $\beta$  genes. A fourth subunit,  $\gamma$ , has been identified in skeletal muscle (Walker, D. et al. (1998) *J. Biol. Chem.* 273:2361-2367; McCleskey, E.W. (1994) *Curr. Opin. Neurobiol.* 4:304-312).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl<sup>-</sup> enters the cell across a basolateral membrane through an  
30 Na<sup>+</sup>, K<sup>+</sup>/Cl<sup>-</sup> cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl<sup>-</sup> from the apical surface, in response to hormonal stimulation, leads to flow of Na<sup>+</sup> and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of

six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) J. Exp. Biol. 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) Curr. Opin. Neurobiol. 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of  $\text{Na}^+$  and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as  $\gamma$ -aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, *supra*). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) Curr. Opin. Neurobiol. 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated  $\text{K}^+$  channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens  $\text{K}^+$  channels to modulate the magnitude of the action potential (Ishi et al., *supra*). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The  $\alpha$  subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated  $\text{K}^+$  channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The  $\beta$  subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, *supra*; Vergara, C. et al. (1998)

Curr. Opin. Neurobiol. 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na<sup>+</sup> channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca<sup>2+</sup> entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an  $\alpha$  subunit which can form functional homomeric channels, and a  $\beta$  subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K<sup>+</sup> channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) Curr. Opin. Neurobiol. 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the G $\beta\gamma$  subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) Curr. Opin. Cell. Biol. 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Brecht (1998) Cell 93:495-498).

### Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) Exp. Nephrol. 4:253-262; Talente, G.M. et al. (1994) Ann. Intern. Med. 120:218-226; and Chillon, M. et al. (1995) New Engl. J. Med. 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle

sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrhythmia disorders such as the long QT syndromes and idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan  
5 (1998) Proc. Natl. Acad. Sci. USA 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) Curr. Opin. Neurology 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) Curr. Opin. Neurobiol. 9:274-280; Cooper, supra).

10 Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) Adv. Pharmacol. 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are  
15 potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na<sup>+</sup> channels have been useful in the treatment of neuropathic pain (Eglen, supra).

Ion channels in the immune system have recently been suggested as targets for  
20 immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious  
25 immunosuppressants (Calahan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, and immunological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of  
30 transporters and ion channels.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4,"



“TRICH-5,” “TRICH-6,” “TRICH-7,” “TRICH-8,” “TRICH-9,” “TRICH-10,” “TRICH-11,”  
“TRICH-12,” “TRICH-13,” “TRICH-14,” “TRICH-15,” “TRICH-16,” “TRICH-17,” “TRICH-18,”  
“TRICH-19,” “TRICH-20,” “TRICH-21,” “TRICH-22,” “TRICH-23,” “TRICH-24,” “TRICH-25,”  
“TRICH-26,” and “TRICH-27.” In one aspect, the invention provides an isolated polypeptide

- 5 comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence  
selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence  
having at least 90% sequence identity to an amino acid sequence selected from the group consisting of  
SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected  
10 from the group consisting of SEQ ID NO:1-27. In one alternative, the invention provides an isolated  
polypeptide comprising the amino acid sequence of SEQ ID NO:1-27.

- The invention further provides an isolated polynucleotide encoding a polypeptide comprising an  
amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the  
group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least  
15 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-  
27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of  
SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-27. In one alternative, the polynucleotide encodes a polypeptide selected  
from the group consisting of SEQ ID NO:1-27. In another alternative, the polynucleotide is selected  
20 from the group consisting of SEQ ID NO:28-54.

- Additionally, the invention provides a recombinant polynucleotide comprising a promoter  
sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid  
sequence selected from the group consisting of a) an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90%  
25 sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c)  
a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID  
NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-27. In one alternative, the invention provides a cell transformed with the  
recombinant polynucleotide. In another alternative, the invention provides a transgenic organism  
30 comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid  
sequence selected from the group consisting of a) an amino acid sequence selected from the group  
consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90%  
sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c)

a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said

target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting

of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:28-54, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for each polypeptide of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of each polypeptide sequence, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of each

polypeptide.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble each polynucleotide sequence, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for each polynucleotide of the invention.

5 Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

10

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will  
15 be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so  
20 forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described.  
25 All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

30 “TRICH” refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other

compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of

TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

5       The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or  
10       synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

      The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to  
15       immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

      The term "antisense" refers to any composition capable of base-pairing with the "sense"  
20       (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense  
25       molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

30       The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.



“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
25	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
30	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
35	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr

Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

---

5

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

10 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one  
 15 biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

20 A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,  
 25 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the  
 30 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:28-54 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:28-54, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:28-54 is useful, for  
 35 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:28-54 from related polynucleotide sequences. The precise length of a fragment of SEQ

ID NO:28-54 and the region of SEQ ID NO:28-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-27 is encoded by a fragment of SEQ ID NO:28-54. A fragment of SEQ ID NO:1-27 comprises a region of unique amino acid sequence that specifically identifies  
5 SEQ ID NO:1-27. For example, a fragment of SEQ ID NO:1-27 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-27. The precise length of a fragment of SEQ ID NO:1-27 and the region of SEQ ID NO:1-27 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

10 A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

15 The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

20 Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

25 For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is  
30 provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other

polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST

5 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

10 *Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

15 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such

20 lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

25 a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment

30 methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment

program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for

annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of

various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical

labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that



hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in

vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, and immunological disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of each of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:28-54 or that distinguish between SEQ ID NO:28-54 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages

comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

5           The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6813453H1 is the identification number of an Incyte cDNA sequence, and ADRETUR01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70207988V1). Alternatively, the identification numbers in column 5 may refer to  
10   GenBank cDNAs or ESTs (e.g., g1947104) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6554406\_006 is the identification number of a Genscan-predicted coding sequence, with g6554406 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding  
15   sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant  
20   with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to  
25   assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural  
30   characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:28-54, as presented in the Sequence Listing, embrace the equivalent RNA sequences,

wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:28-54.

Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of

hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:28-54 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in

5 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or

10 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system

15 (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

20 The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)

25 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al.

30 (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo

Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number



5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp.55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences

encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.)

The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc.*

Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *apr*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing

sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

5 In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

10 Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding  
15 assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and  
20 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and  
25 may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as  
30 well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing

polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds

that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES)

cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest  
5 disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).  
10 Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from  
15 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create “knockin” humanized animals  
20 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a  
25 mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. Therefore, TRICH appears to play a  
30 role in transport, neurological, muscle, and immunological disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be



administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,

postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); and an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such

disorders include, but are not limited to, those transport, neurological, muscle, and immunological disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

5 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate  
10 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

15 An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and  
20 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to  
25 increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to  
30 TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J.*

5 *Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc.*  
10 *Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton,  
15 D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

20 Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al.  
25 (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its  
30 specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association

constant,  $K_a$ , which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other

gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

5 In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency  
10 (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii)  
15 express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides  
20 brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in  
25 TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-  
30 217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF,

PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. U.S.A. 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene

therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. U.S.A. 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver  
5 polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are  
10 described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver  
15 polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has  
20 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under  
25 the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids  
30 containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus,



Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for

secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by  
5 any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that  
10 synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages  
15 within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a  
20 compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide  
25 sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes  
30 expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary

library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample  
5 may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus  
10 forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression  
15 system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide  
20 sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

25 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and  
30 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's

Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose

lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be

quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:28-54 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral

neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis,

5 infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia,

10 hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's

15 disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and

20 Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

25 neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration,

30 and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's

35 syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and

myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); and an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with



values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

5           Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

10           With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further  
15           progression of the cancer.

          Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and  
20           will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

          In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are  
25           substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal  
30           tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP

(isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the  
5 alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et  
10 al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the  
15 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor  
20 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

25 In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of  
30 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

5 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of  
10 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a  
15 signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the  
20 rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released  
25 February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated  
30 biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are

5 analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl

10 sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or

15 untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In

20 some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111;

25 Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation

30 between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference  
5 in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated  
10 with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,  
15 Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999)  
20 Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members  
25 of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat.  
30 Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical  
5 map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may  
10 reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g.,  
15 Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug  
20 screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT  
25 application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a  
30 solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

5 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/172,000, U.S. Ser. No. 60/176,083, U.S. Ser. No. 60/177,332, U.S. Ser. No. 60/178,572, U.S. Ser. No. 60/179,758, and U.S. Ser. No. 60/181,625, are expressly incorporated by reference herein.

## EXAMPLES

### 15 I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

20 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

30 In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the

appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

5 PBLUESCRIPT plasmid (Stratagene), PSFORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

10 Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid

15 purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-

20 well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

25 Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI

30 PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA



sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families.

See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMR. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length.

Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA.

The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences which were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and

polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:28-54. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

5 Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an  
10 assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were  
15 also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or  
20 public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were  
25 derived entirely from edited or unedited Genscan-predicted coding sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped  
30 to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in

the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of TRICH Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:28-54 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:28-54 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment

of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter

of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### **VIII. Extension of TRICH Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer

pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; 5 Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II 10 (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and 15 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, 20 and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 25 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC 30 DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

## IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:28-54 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments.

- 5 Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per
- 10 minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

- The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16
- 15 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

## X. Microarrays

- 20 The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure
- 25 analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol.
- 30 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the

biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of

- 5 complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

- Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with
- 15 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.
- 20 (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

#### **Microarray Preparation**

- 25 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia
- 30 Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and



coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## **XI. Complementary Polynucleotides**

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

## **XII. Expression of TRICH**

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory

element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XIV. Production of TRICH Specific Antibodies**

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH

activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring TRICH Using Specific Antibodies**

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

#### **XVI. Identification of Molecules Which Interact with TRICH**

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as G $\beta$ y proteins (Reimann, supra) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, supra). TRICH, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the reconstitutes a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH

ion channel activity using the assays described in section XVIII.

#### XVII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as  $\beta$ -galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and  $\beta$ -galactosidase.

Transformed cells expressing  $\beta$ -galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or  $\beta$ -galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18°C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or  $\text{Ca}^{+2}$  (in the form of  $\text{CaCl}_2$ ), where appropriate. Electrode resistance is set at 2-5 M $\Omega$  and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per

oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM CaCl<sub>2</sub>, 1mM MgCl<sub>2</sub>, 1mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM Hepes, 3.8 mM NaOH, 50µg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM CaCl<sub>2</sub>, 1mM MgCl<sub>2</sub>, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with <sup>3</sup>H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na<sup>+</sup>-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP-[γ-<sup>32</sup>P], separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered <sup>32</sup>P using a scintillation counter. The reaction mixture contains ATP-[γ-<sup>32</sup>P] and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of <sup>32</sup>P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

#### XVIII. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the Ca<sup>2+</sup> indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl<sup>-</sup> indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonyl dyes such as DiBAC<sub>4</sub> (Molecular Probes). DiBAC<sub>4</sub> equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of DiBAC<sub>4</sub> entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be

selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will  
5 be apparent to those skilled in the art without departing from the scope and spirit of the invention.  
Although the invention has been described in connection with certain embodiments, it should be  
understood that the invention as claimed should not be unduly limited to such specific embodiments.  
Indeed, various modifications of the described modes for carrying out the invention which are obvious  
to those skilled in molecular biology or related fields are intended to be within the scope of the following  
10 claims.



Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1416107	1	1416107CD1	28	1416107CBI
1682513	2	1682513CD1	29	1682513CBI
2446438	3	2446438CD1	30	2446438CBI
2817822	4	2817822CD1	31	2817822CBI
4009329	5	4009329CD1	32	4009329CBI
6618083	6	6618083CD1	33	6618083CBI
7472002	7	7472002CD1	34	7472002CBI
1812692	8	1812692CD1	35	1812692CBI
3232992	9	3232992CD1	36	3232992CBI
3358383	10	3358383CD1	37	3358383CBI
4250091	11	4250091CD1	38	4250091CBI
70064803	12	70064803CD1	39	70064803CBI
70356768	13	70356768CD1	40	70356768CBI
5674114	14	5674114CD1	41	5674114CBI
1254635	15	1254635CD1	42	1254635CBI
1670595	16	1670595CD1	43	1670595CBI
1859560	17	1859560CD1	44	1859560CBI
5530164	18	5530164CD1	45	5530164CBI
139115	19	139115CD1	46	139115CBI
1702940	20	1702940CD1	47	1702940CBI
1703342	21	1703342CD1	48	1703342CBI
1727529	22	1727529CD1	49	1727529CBI
2289333	23	2289333CD1	50	2289333CBI
2720354	24	2720354CD1	51	2720354CBI
3038193	25	3038193CD1	52	3038193CBI
3460979	26	3460979CD1	53	3460979CBI
7472200	27	7472200CD1	54	7472200CBI

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1416107CD1	g7018605	1.9e-302	Glucose transporter [Rattus norvegicus] (Ibberson, M. et al. (2000) J. Biol. Chem. 275:4607-4612)
2	1682513CD1	g5263196	1.4e-153	Stretch-inhibitable nonselective channel (SIC) [Rattus norvegicus] (Cloning of a stretch-inhibitable nonselective cation channel. J Biol Chem. 1999 Mar 5;274(10):6330-6335)
3	2446438CD1	g4589141	0	Vanilloid receptor-like protein 1 [Homo sapiens] (A capsaicin-receptor homologue with a high threshold for noxious heat. Nature 1999 398:436-441)
5	4009329CD1	g3873983	1.9e-64	Similar to Na+/Ca+, K+ antiporter [C. elegans]
6	6618083CD1	g9230651	4.7e-268	Facilitative glucose transporter family member GLUT9 [Homo sapiens] (Phay, J.E. et al. (2000) Genomics 66:217-220)
7	7472002CD1	g433960	0	Aorta CNG channel (rACNG) [Oryctolagus cuniculus] (Primary structure and functional expression of a cyclic nucleotide-gated channel from rabbit aorta. FEBS Lett. 1993 Aug 23;329(1-2):134-138)
8	1812692CD1	g3928756	4.5e-48	Transient receptor potential channel 7 [Homo sapiens] (Nagamine, K. et al. (1998) Molecular cloning of a novel putative Ca2+ channel protein (TRPC7) highly expressed in brain)
9	3232992CD1	g3874275	3.2e-70	Similarity to yeast low-affinity glucose transporter HXT4 [Caenorhabditis elegans]
10	3358383CD1	g3004482	1.4e-163	Putative integral membrane transport protein [Rattus norvegicus] (Schomig, E. et al. (1998) Molecular cloning and characterization of two novel transport proteins from rat kidney. FEBS Lett. 425:79-86)
11	4250091CD1	g3880445	5.7e-16	VM106R.1 (similar to K+ channel tetramerisation domain) [Caenorhabditis elegans]
12	70064803CD1	g3874275	7.0e-84	Similarity to yeast low-affinity glucose transporter HXT4 [Caenorhabditis elegans]
13	70356768CD1	g183298	4.1e-54	GLUT5 protein [Homo sapiens] (Kayano, Y. et al. (1990) Human facilitative glucose transporters. Isolation, functional characterization, and gene localization of cDNAs encoding an isoform (GLUT5) expressed in small intestine, kidney, muscle, and adipose tissue and an unusual glucose transporter pseudogene-like sequence (GLUT6). J. Biol. Chem. 265:13276-13282)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
14	5674114CD1	g5771352	1.3e-238	Inward rectifier potassium channel Kir2.4 [Homo sapiens] (Topert, C. et al. (1998) Kir2.4: a novel K <sup>+</sup> inward rectifier channel associated with motoneurons of cranial nerve nuclei. J. Neurosci. 18:4096-4105)
15	1254635CD1	g3953533	1.7e-210	Inwardly rectifying potassium channel Kir5.1 [Mus musculus] (Mouri, T. et al. (1998) Assignment of mouse inwardly rectifying potassium channel Kcnj16 to the distal region of mouse chromosome 11. Genomics 54:181-182)
16	1670595CD1	g9502260	2.3e-146	Cation-chloride cotransporter-interacting protein [Homo sapiens] (Caron, L. et al. (2000) J. Biol. Chem. 275:32027-32036)
17	1859560CD1	g5834394	1.4e-101	Sulfate transporter [Drosophila melanogaster]
18	5530164CD1	g4903004	3.5e-20	UDP-N-acetylglucosamine transporter [Homo sapiens] (Ishida, N. et al. (1999) Molecular cloning and functional expression of the human golgi UDP-N-acetylglucosamine transporter. J. Biochem. 126:68-77.)
19	139115CD1	g8131858	1.5e-49	Putative thymic stromal co-transporter TSCOT [Mus musculus] (Kim, M.G. et al. (2000) J. Immunol. 164:3185-3192)
20	1702940CD1	g5725224	2.5e-143	bK212A2.2 (similar to apolipoprotein L) [Homo sapiens]
21	1703342CD1	g6003536	8.1e-06	Calcium channel alpha-1 subunit [Bdelloura candida]
22	1727529CD1	g4529890	0.0	NG22 [Homo sapiens]
23	2289333CD1	g4539333	5.6e-35	Putative amino acid transport protein [Arabidopsis thaliana]
24	2720354CD1	g3875242	4.3e-38	Similar to mitochondrial carrier protein [Caenorhabditis elegans]
26	3460979CD1	g1931644	4.7e-08	Membrane protein PTM1 precursor isolog (putative major facilitator superfamily transporter) [Arabidopsis thaliana]
27	7472200CD1	g2811254	2.8e-21	Amiloride-sensitive Na <sup>+</sup> channel [Drosophila melanogaster] (Adams, C.M. et al. (1998) J. Cell Biol. 140:143-152)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1416107CD1	477	S99 T2 T281 S430 T205	N349	Sugar transporter domain: A29-F474 Sugar transport protein signatures: V108-L174, L293-S350, G41-I51, L124-V143, Q267-F277, V375-L396, S398-T410 Glucose transporter signatures: F257-Y278, V375-S398, G439-V459 Transmembrane domains: I259-A279, L293-M313, L320-Y339, Y438-F457	MOTIFS HMMER-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS
2	1682513CD1	498	S47 T131 S286 T367 S463 T67 S315 S382	N278 N411 N429	Glucose transporter signature: P319-N339 Transmembrane domains: A95-Y117, V142-F160, L178-Y201, A200-F219, F244-L263, P319-N339	MOTIFS BLIMPS- PRINTS HMMER
3	2446438CD1	764	T64 T329 S23 S67 T106 S268 S339 S348 S353 S464 S468 S667 T692 S697 T720 T101 T115 S325 T414 Y110 Y227 Y333	N570	Ankyrin repeat: R162-C194; F208-S243 Q293-F328 Transmembrane domains: L386-F405, I463-V486, F538-S557, L623-I642	MOTIFS HMMER-PFAM HMMER
4	2817822CD1	255	T30 S167 T33 T71 S23 T50 S134 T162 T238		Potassium channel signature: R76-T95	MOTIFS BLIMPS- PRINTS
5	4009329CD1	584	S258 S321 T70 S271 S273 S468 S514 S62 T132	N60 N125	Signal peptide: M1-G29 Sodium/calcium exchanger protein domain: I113-Q252, L431-F576 Transmembrane domains: T101-F121, T166-I189, L234-Y251, Y382-A402, F492-R513, L560-M584	MOTIFS HMMER HMMER-PFAM HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	6618083CD1	416	T111 S4 S164 S274 T374 S16 S226 S269	N45 N61 N410	Signal peptide: M1-G37 Sugar transporter domain: A30-E416 Sugar transport proteins signatures: A123-L189, S39-V49, I139-M158, Y298-F308 Glucose transporter signatures: V288-Y309, I356-Q376 Transmembrane domains: V83-V99, I356-L375	MOTIFS SPScan HMMER HMMER-PFAM BLIMPS-BLOCKS ProfileScan BLIMPS-PRINTS HMMER
7	7472002CD1	664	S402 S40 S46 S93 T107 T313 T337 T381 T422 S476 S552 T591 T606 T634 T2 S35 T124 T208 T418 T448 Y648	N311 N379	Transmembrane region, cyclic nucleotide gated channel: Y215-I440 Cyclic nucleotide binding domain: K469-D565, A460-S476, G478-V501, G516-L525 Transmembrane domain: Y350-I375	MOTIFS HMMER-PFAM BLIMPS-BLOCKS HMMER
8	1812692CD1	242	T95 S35 S37 T124 S204 S221 S2 S9 S20 T21 S86 Y36	N15 N84	Protein melastatin chromosome transmembrane PD018035: Y117-I227	BLAST- PRODOM
9	3232992CD1	398	T307 S338 S100 T133 T241 T303 S377 S395	N161	Transmembrane domains: A217-Q242, L247-F264, L350-F368 Sugar transport proteins signature: L45-G94, V30-I96	HMMER MOTIFS BLIMPS-BLOCKS ProfileScan
10	3358383CD1	553	S337 S352 S409 T58 S60 S109 T133 S337 T433 T527 S167 S201 T226 S282 T323 T405	N39 N56 N62 N102 N107 N473	Transmembrane domains: F204-A222, M470-Y493, I500-T519 glpT family of transporters: V151-D168 Organic transport protein, renal anion transporter, cationic kidney specific solute PD151320: N102-L144	HMMER MOTIFS BLIMPS-BLOCKS BLAST- PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	4250091CD1	213	S2 S93 S172 S184 T17 T22 T137 S210 Y89	N188	Potassium channel signature: Q48-T67	MOTIFS BLIMPS- PRINTS
12	70064803CD1	476	T365 S11 S364 S453 S292 T361 S390 S466		Transmembrane domains: V222-R239, G327-V350, M413-F432 Sugar transport proteins signature: L153-G202, V138-I204	HMMER  MOTIFS BLIMPS- BLOCKS ProfileScan
13	70356768CD1	246	S100 S238 S118 S215	N34 N50	Signal peptide: M1-G27 Sugar transport proteins signature: I127-G176, A112-V178, T28-I38, M128-M147, M133-R158	MOTIFS SPScan HMMER BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS
14	5674114CD1	436	S11 T80 S154 S340 S362 T263 S376 S422 Y47	N195	Transmembrane domain: M163-L181	HMMER
					Sugar transport proteins: DM00135 P46408 : A112-C229	BLAST-DOMO
					Inward rectifier potassium channel domain: V53-L394, R72-L118, P126-Q169, C170-V199, A204-Q238, D296-Y346, T358-E368	MOTIFS HMMER-PFAM BLIMPS-PFAM
					Transmembrane domain: W88-L114 Inward rectifier potassium channel: DM00448 P52188 : N34-A395	HMMER BLAST-DOMO
					Inward rectifier potassium channel: PD001103: V53-Q372 KIR2.4 protein: PD124342: A373-P436 PD063376: M1-F52	BLAST- PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	1254635CD1	453	S408 T16 T99 S416 S29 T209 T216 T250 S375		Signal peptide: M1-A32 Transmembrane domains: F109-A131, V182-I200 Inward rectifier potassium channel domain: L72-T399 Inward rectifier potassium ion channel subfamily PD001103: K74-K403 Voltage-gated inward rectifier potassium channel BIR9, KIR5.1, transmembrane PD063375: M36-H73 Inward rectifier potassium channel: DM00448 P52185 27-380: K64-L379 Inward rectifier potassium channel signature: S91-L137, P145-Q188, S189-R218, A223-R257, N310-C360, A371-W381 Transmembrane domains: L37-L58, A74-I93, L134-Y154, F159-V179 Sensitive cotransporter, chloride, sodium: DM01178 S06903 1-128: A32-L153	SPScan HMMER HMMER-PFAM BLAST- PRODOM BLAST-DOMO BLIMPS-PFAM
16	1670595CD1	299	S38 T211 S263 S33 T187	N193 N208		HMMER BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	1859560CD1	606	T96 T116 S298 S571 T572 S595 T245	N294	Transmembrane domains: L45-I63, T396-T421 Sulfate transporter family domain: M137-A468 Sulfate transporters protein signature: A54-I107, L125-L176 Sulfate transport protein, transmembrane, permease PD001255: M137-L465 Sulfate transport protein, transmembrane, permease PD001121: L30-G143 Sulfate transporter: DM01229 S64926 69-531: L30-W428 Sulfate transporters motif: P77-R98	HMMER HMMER-PFAM BLIMPS-BLOCKS BLAST-PRODROM BLAST-PRODROM BLAST-DOMO MOTIFS
18	5530164CD1	324	S2 S139 Y115	N99 N100 N101 N232	Transmembrane domains: A145-V163, L302-L319	HMMER
19	1391115CD1	445	S54 S32 S77 S217 S424 S438 T150 S237 S443 Y23	N22 N30 N37 N127 N213 N235	Transmembrane domains: W182-F200, F242-I261, Y283-F302 Sugar transporter motif: L75-S91 Glucose transporter signature: W182-L202	HMMER MOTIFS BLIMPS-PRINTS
20	1702940CD1	337	T30 S167 T208 S306		Apolipoprotein L precursor, lipid transport, glycoprotein, signal, DJ6802.1 PD042084: M1-D336	BLAST-PRODROM
21	1703342CD1	273	T3 S63 T222 S248 S250 T10 S98 S219 S224		Transmembrane domains: F101-A118, M142-F161, F170-I186, I202-I218 Ion transport protein domain: L95-L269 (Score: -132.1, E-value: 0.72)	HMMER HMMER-PFAM



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22	1727529CD1	710	S31 S102 S119 T135 S304 S22 S218 S430 S431 T494 S573 S619 Y13	N29 N69 N155 N197 N298 N393 N405 N416 N678	Transmembrane domains: C38-Y58, V241-L266, W309-V326, F356-T375, F440-L458, T499-I522, L598-F618, I645-V663 ABC 3 transport family: S228-Q427 (Score: -182.9, E-value: 2.1) Anion exchanger signature: A311-L330 Transmembrane domains: L54-I81, V127-F145, Y184-S208, L279-G297, I331-K357, I426-T451 Transmembrane amino acid transporter protein domain: A55-F436 Amino acid transporter protein, permease, transmembrane, putative proline PD001875: D27-I337 Signal peptide: M1-G15 Mitochondrial carrier proteins domain: S25-L109, L122-T202 Mitochondrial energy transfer proteins signature: L128-Q152 Mitochondrial energy transfer proteins signature: L27-I75, V123-Q171 Mitochondrial carrier proteins signature: G87-D107, V136-D154 Adenine nucleotide translocator 1 signature: R63-V84, E176-R191 Mitochondrial carrier proteins motifs: P46-L54, P143-L151 Transport protein, transmembrane, inner mitochondrial, ADP/ATP: PD000117: L31-F200 Mitochondrial energy transfer protein: DM00026 P38087 243-325: L128-Y209	HMME HMME-PFAM BLIMPS- PRODOM HMME HMME-PFAM BLAST- PRODOM SPScan HMME-PFAM BLIMPS- BLOCKS ProfileScan BLIMPS- PRINTS BLIMPS- PRINTS MOTIFS BLAST- PRODOM BLAST-DOMO
23	2289333CD1	476	T97 T7 S8 S125 T443 S272 S322 T351 T451 Y184	N166 N169 N212 N425 N467		
24	2720354CD1	237	T17 T64 S172			

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	3038193CD1	345	T204 T251 S57 S243 T263 T308 S340	N246	Transmembrane domains: L67-L95, I134-I156, I224-F242 Sodium bile acid symporter family: Y44-P212 (Score: -7.0, E-value: 9.0e-4) Phosphate transporter signature: F153-G171	HMMER  HMMER-PFAM
26	3460979CD1	521	S115 T184 S75 T93 S100 S126 S128 S134 S148 S183 S213 S256 S363 S389 S430 S510 T171 S180 S235 T247 S422 Y506	N70 N169 N211	Transmembrane domains: L265-L284, I335-I361 Protein precursor PTM1, transmembrane, signal PD014374: G219-E517 (P-value: 7.1e-07)	BLIMPS- PRODOM  HMMER  BLAST- PRODOM
27	7472200CD1	555	T43 S56 T92 T148 T298 S423 S468 S20 S52 S82 S96 T184 S208 S252 S393	N132 N175 N311 N361 N421	Amiloride-sensitive sodium channel alpha subunit signature PR01078: Y102-N118, Y342-Q353, Q353-P370, Q388-N404, G455-E471 Transmembrane domain: V452-F475 Amiloride-sensitive sodium channel ASC: F38-L476 Amiloride-sensitive sodium channel proteins BL01206: R37-L47, Y342-F368, L427-L472	BLIMPS- PRINTS  HMMER HMMER-PFAM  BLIMPS- BLOCKS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
28	1416107CB1	2080	1-109, 1901-2080, 1363-1446	g1941704	116	609
				6813453H1 (ADRETUR01)	319	870
				6605280H1 (UTREDIT07)	820	1447
				881845R1 (THYRNOT02)	889	1479
				1416107F6 (BRAINOT12)	1348	1896
				1416107T6 (BRAINOT12)	1579	2080
				71826604V1	1563	1974
				7448905T1 (BRAYDIN03)	1578	2050
				6300413H1 (UTREDIT07)	423	752
				71805807V1	750	1580
				71827149V1	1584	2080
				71807187V1	701	1241
				6813453R6 (ADRETUR01)	1	312
				70207988V1	1	469
29	1682513CB1	2128	1-1535, 1560-1581	70213506V1	394	872
				70211216V1	489	985
				70210573V2	852	1468
				70207907V1	988	1512
				70210540V2	1357	1948
				2866122T6 (KIDNNOT20)	1548	2108
				70211461V1	1597	2128
				5073532H2 (COLCTUT03)	1	311
				6309494H1 (NERDTN03)	260	812
				6268005H1 (MCLDTXN03)	344	996
30	2446438CB1	2825	1-65, 2000-2202, 999-1820	70382927D1	996	1525
				70386205D1	1469	2075
				1798255F6 (COLNNOT27)	1632	2194
				1562088F6 (SPLNNOT04)	2178	2727
				2514370F6 (LIVRTUT04)	2303	2825
				1502510F6 (BRAITUT07)	1	439
				70271734V1	183	768
				70273052V1	431	930
				70271651V1	891	1453
				2817822F6 (BRSTNOT14)	981	1538
				70272460V1	1094	1718
31	2817822CB1	1718	1-71, 609-914			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
32	4009329CB1	2000	1-962	6456193H1 (PLACFEB01)	1	640
				6780428J1 (OVARDIR01)	582	1260
				6307863H1 (NERDTDN03)	725	1364
				6781250H1 (OVARDIR01)	972	1639
				7253109J1 (PROSTME05)	1514	1842
33	6618083CB1	2216	1-96, 1201-2216	6759035J1 (HEAONOR01)	1515	2000
				5722362H1 (SEMNOT05)	1	581
				70789558V1	504	1127
				70787652V1	588	1203
				70791819V1	1050	1650
34	7472002CB1	1995	1-862, 1766-1995	70787819V1	1361	1984
				70791126V1	1695	2216
				g2121300.v113.gs_2.nt.edit	1	1995
				1812692F6 (PROSTUT12)	564	984
				5425924F6 (PROSTMT07)	1	488
35	1812692CB1	988	1-147, 244-570	g2525933	823	988
				5000833F6 (PROSTUT21)	283	804
				224000R6 (PANCNOT01)	2435	3087
				6825934J1 (SINTNOR01)	1	515
				7062063H1 (PENITMN02)	2683	3179
36	3232992CB1	3179	2106-2665, 1-1646	4491105H1 (BRAMDI02)	2167	2861
				1698347F6 (BLADTUT05)	1762	2333
				70053653D1	1392	1870
				1807402F6 (SINTNOR13)	476	1019
				70055908D1	1317	1837
37	3358383CB1	1986	1465-1986, 1340-1371	7170824H2 (BRSTTMC01)	719	1373
				6555265H1 (BRAFNOR02)	1859	2372
				g1444660	992	1464
				g1009986	768	1254
				027195T6 (SPLNFET01)	1674	1986
				g1505781	552	1252
				3358383T6 (PROSTUT16)	1341	1691
				6221856U1	585	1252
				6221857U1	1	727

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
38	4250091CB1	3294	1-920, 1991-2034, 2488-2760, 1365-1630	g715570	2893	3294
				70759966V1	1864	2477
				4250091F6 (BRADDIR01)	1	532
				5715843H1 (PANCNOT16)	2539	3235
				70789723V1	444	1032
				966456R6 (BRSTNOT05)	2862	3293
39	70064803CB1	2043	1-22, 544-1285	70759467V1	1199	1842
				70788682V1	604	1302
				7056848H1 (BRALNON02)	2373	2989
				858645R1 (BRAITUT03)	1940	2507
				70761829V1	1333	1947
				2758549R6 (THPLAZS08)	1540	2043
40	70356768CB1	1915	1241-1263, 853-897, 1-143, 1450-1532, 668-820	6810024J1 (SKIRNOR01)	576	1280
				1676182T6 (BLADNOT05)	1361	2019
				2109762R6 (BRAITUT03)	1181	1797
				70503885V1	501	1183
				7177480H2 (BRAXDIC01)	1	534
				70450108V1	509	1081
41	5674114CB1	1809	1-402	70451575V1	1072	1730
				1468307F6 (PANCUT02)	1	524
				70451567V1	368	1078
				70449058V1	1384	1915
				70449392V1	1060	1720
				6776218J1 (OVARDIR01)	1078	1809
42	1254635CB1	1730	1-106, 1711-1730, 567-635, 696-900	3024042H1 (PROSDIN01)	690	1040
				6292787H1 (BMAUNA01)	939	1290
				6776218H1 (OVARDIR01)	184	944
				g5686663.v113.gs.i6.nt	1	1311
				2613664F6 (ESOGTUT02)	655	1177
				SXBC01035V1	75	573
43	1670595CB1	1147	746-980, 1-696	2863343F6 (KIDNNOT20)	1	515
				2614317T6 (GBLANOT01)	1126	1730
				SCSA01493V1	548	724
				3323244T6 (PTHYNOT03)	960	1627
				SCIA02891V1	368	1147
				SCIA04658V1	1	641

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
44	1859560CB1	2745	1-820, 865-2059	6195927H1 (PITUNO01)	2367	2745
				824186R1 (PROSN006)	1159	1718
				1399644F6 (BRAITUT08)	503	973
				6812696J1 (ADRETUR01)	30	732
				7255024H1 (FIBRTXC01)	814	1364
				2127239R7 (KIDNNOT05)	1732	2176
				5990868H1 (FTUBTUT02)	1493	1795
				6826978H1 (SINTNOR01)	1	479
				6850265H1 (BRAIFEN08)	2056	2740
				4677711H1 (NOSEDIT02)	1107	1381
				1859560T6 (PROSN018)	2247	2745
				4320381H1 (BRADDT02)	1925	2202
				7175713H1 (BRSTMC01)	178	811
45	5530164CB1	3204	1241-2064, 1-548, 572-1097	3217236H1 (TESTNOT07)	530	812
				2552002H1 (LUNGTUT06)	1	242
				70039789V1	2452	3175
				70090155V1	1781	2492
				6830248J1 (SINTNOR01)	542	1211
				6729730H1 (COLITUT02)	1214	1889
				2850920F6 (BRSTTUT13)	929	1416
				6125934H1 (BRAHNON05)	2505	3204
				6059181H1 (BRAENOT04)	1923	2496
				1956752F6 (CONNNOT01)	1452	1893
				6455739H1 (COLNDIC01)	528	1219
				70077060U1	1900	2535
				7126228H1 (COLNDIY01)	1	580
46	139115CB1	2763	1-770, 1345-1389, 1455-1683	2468105F6 (THYRNOT08)	827	1463
				70079483U1	2145	2763
				70122236V1	1424	1971
				70122363V1	1451	1976
				351595R1 (LVENNOT01)	2066	2566
				70480521V1	1027	1639
				4335403F6 (KIDCTMT01)	1	516
				70466195V1	559	1157
				70466476V1	494	1102
47	1702940CB1	1639	1-246, 1536-1639			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
48	1703342CB1	1600	1-812	3348562H1 (BRAITUT24)	1	282
				285125R1 (EOSIHET02)	1041	1598
				7071066H1 (BRAUTDR02)	250	854
				6494627H1 (BONRNOT01)	1220	1600
				6879086J1 (LNODNOR03)	360	1094
49	1727529CB1	2380	1-569, 1228-1654	957891H1 (KIDNNOT05)	2085	2380
				60211961U1	691	1237
				6800135J1 (COLENOT03)	1250	1983
				6798918J1 (COLENOT03)	1693	2284
				60211964U1	262	807
				6798894H1 (COLENOT03)	1089	1774
				3249035F6 (SEMVNOT03)	1	626
				3566495H1 (BRONNOT02)	1984	2298
				2552315T6 (LUNGTTUT06)	1322	1862
				9872898	848	1328
50	2289333CB1	3038	1-611, 2497-2524	1435329F1 (PANCNOT08)	2197	2725
				3553901H1 (SYNONOT01)	2570	2865
				2508452H1 (CONUTUT01)	1	114
				2771704H1 (COLANOT02)	1815	2078
				6999443H1 (HEALDIR01)	2	553
				91665184	2594	3038
				2289333R6 (BRAINON01)	1254	1708
				91156003	2524	3032
				5597992H1 (UTRENON03)	1029	1286
				95545742	612	1066
				5836345H1 (BRAIDIT05)	2624	2880
				4220788F6 (PANCNOT07)	613	958
				1994713T6 (BRSTTUT03)	1949	2451
				2040880R6 (HIPONON02)	463	821
				2720354F6 (LUNGTTUT10)	490	1046
				6942433H1 (FTUBTUR01)	885	1437
				6121303H1 (BRAHNON05)	1630	2340
				6558224H1 (BRAFNON02)	1713	2406
				6940932H1 (FTUBTUR01)	1	465
51	2720354CB1	2608	1-2058	91927466	325	872
				6826181J1 (SINTNOR01)	1062	1666
				6197805H1 (PITUNON01)	2154	2608

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragments	Sequence Fragments	5' Position	3' Position
52	3038193CB1	3804	3392-3457, 1169-1264, 1-829, 2271-2483, 1319-1363	044564H1 (TBLYN0T01)	2311	2562
				901446R6 (BRSTTUT03)	1733	2282
				g4088232	3459	3804
				1428831H1 (SINTBST01)	473	661
				2741328T6 (BRSTTUT14)	3235	3804
				4970206H1 (KIDEUNC10)	1029	1303
				2768967H1 (COLANOT02)	772	1020
				5688762F6 (BRAIUNT01)	36	621
				3038193F6 (BRSTNOT16)	1284	1710
				6477440H1 (PROSTMC01)	1843	2486
				70809191V1	2411	2832
				3154867H1 (TLYMTXT02)	1	272
				2257401R6 (OVARTUT01)	2660	3167
				g4268882	1421	1815
				2257401T6 (OVARTUT01)	2896	3520
53	3460979CB1	1894	1-36 1746-1894	2237852F6 (PANCUT02)	519	945
				3460979F6 (293TF1T01)	1000	1500
				7161336H1 (PLACNOR01)	582	1193
				6800921J1 (COLENOR03)	1	557
54	7472200CB1	1668	1-1668	7057496H1 (BRAIINON02)	1206	1894
				GNN.g6554406_006	1	1668



Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
28	1416107CB1	UTREDIT07
29	1682513CB1	SPLNNOT11
30	2446438CB1	MCLDTXN03
31	2817822CB1	BRAITUT07
32	4009329CB1	OVARDIN02
33	6618083CB1	HELAUNT01
35	1812692CB1	PROSTUT12
36	3232992CB1	PANCN0T01
37	3358383CB1	PROSTUT16
38	4250091CB1	BRAITUT03
39	70064803CB1	THP1AZS08
40	70356768CB1	HNT2AGT01
41	5674114CB1	OVARDIR01
42	1254635CB1	LUNGFET03
43	1670595CB1	BRAITUT24
44	1859560CB1	NGANN0T01
45	5530164CB1	BRAYDIN03
46	139115CB1	SINTNOT18
47	1702940CB1	BRAVXTX04
48	1703342CB1	EOSIHET02
49	1727529CB1	PROSN0T18
50	2289333CB1	LUNGTUT06
51	2720354CB1	PROSTUS23
52	3038193CB1	LIVRN0N08
53	3460979CB1	COLENOR03

Table 6

Library	Vector	Library Description
UTREDIT07	pINCY	Library was constructed using RNA isolated from diseased endometrial tissue removed from a female during endometrial biopsy. Pathology indicated in phase endometrium with missing beta 3, Type II defects.
SPLNNOT11	pINCY	Library was constructed using RNA isolated from diseased spleen tissue removed from a 14-year-old Asian male during a total splenectomy. Pathology indicated changes consistent with idiopathic thrombocytopenic purpura. The patient presented with bruising. Patient medications included Vincristine.
MCLDTXN03	pINCY	Library was constructed from a pool of two dendritic cell libraries. Starting libraries were constructed using RNA isolated from untreated and treated dendritic cells from umbilical cord blood CD34+ precursor cells removed from a male. The cells were derived with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), and stem cell factor (SCF). The libraries were normalized under conditions adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228 and Bonaldo et al. (1996) Genome Res. 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAITUT07	pINCY	Library was constructed using RNA isolated from left frontal lobe tumor tissue removed from the brain of a 32-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated low grade desmoplastic neuronal neoplasm. The patient presented with nausea, vomiting, and headache. Patient history included alcohol, tobacco use, and marijuana use twice a week for six years. Family history included atherosclerotic coronary artery disease in the grandparent(s).
OVARDIN02	pINCY	Library was constructed from an ovarian tissue library. Starting RNA was made from diseased ovarian tissue removed from a 39-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, dilation and curettage, partial colectomy, incidental appendectomy, and temporary colostomy. Pathology indicated the right and left adnexa, mesentery and muscularis propria of the sigmoid colon were extensively involved by endometriosis. Endometriosis also involved the anterior and posterior serosal surfaces of the uterus and the cul-de-sac. The endometrium was proliferative. Pathology for the associated tumor tissue indicated multiple (3 intramural, 1 subserosal) leiomyomata. The patient presented with abdominal pain and infertility. Patient history included scoliosis. Previous surgeries included laparoscopic cholecystectomy and exploratory laparotomy. Patient medications included Megace, Danazol, and Lupron. Family history included hyperlipidemia in the mother, benign hypertension, hyperlipidemia, atherosclerotic coronary artery disease, coronary artery bypass graft, depressive disorder, brain cancer, and type II diabetes. The library was normalized under conditions adapted from Soares et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228 and Bonaldo et al. (1996) Genome Res. 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
HELAUNT01	pINCY	Library was constructed from RNA isolated from an untreated HeLa cell line, derived from cervical adenocarcinoma removed from a 31-year-old Black female.

Table 6 (cont.)

Library	Vector	Library Description
BRAITUT03	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 17-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a grade 4 fibrillary giant and small-cell astrocytoma. Family history included benign hypertension and cerebrovascular disease.
HNT2AGT01	PBLUESCRIPT	Library was constructed at Stratagene (STR37233), using RNA isolated from the hNT2 cell line derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor. Cells were treated with retinoic acid for 5 weeks and with mitotic inhibitors for two weeks and allowed to mature for an additional 4 weeks in conditioned medium.
OVARDIR01	pcDNA2.1	Library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex.
PANCN0T01	PBLUESCRIPT	Library was constructed using RNA isolated from the pancreatic tissue of a 29-year-old Caucasian male who died from head trauma.
PROSTUT12	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
PROSTUT16	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 55-year-old Caucasian male. Pathology indicated adenocarcinoma, Gleason grade 5+4. Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Patient history included calculus of the kidney. Family history included lung cancer and breast cancer.
THPLAZS08	PSPORT1	This subtracted THP-1 promonocyte cell line library was constructed using 5.76 x 1e6 clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The hybridization probe for subtraction was derived from a similarly constructed library, made from RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954, and Bonaldo et al., Genome Research (1996) 6:791. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer 26 (1980):171).

Table 6 (cont.)

Library	Vector	Library Description
BRAITUT24	pINCY	Library was constructed using RNA isolated from right frontal brain tumor tissue removed from a 50-year-old Caucasian male during a cerebral meninges lesion excision. Pathology indicated meningioma. Family history included colon cancer and cerebrovascular disease.
BRAYDIN03	pINCY	This normalized brain tissue library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones.
LUNGFET03	pINCY	Library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
NGANNOT01	PSPORT1	Library was constructed using RNA isolated from tumorous neuroganglion tissue removed from a 9-year-old Caucasian male during a soft tissue excision of the chest wall. Pathology indicated a ganglioneuroma. Family history included asthma.
BRAVXT04	PSPORT1	Library was constructed using RNA isolated from separate populations of human astrocytes stimulated for 4 to 6 hours with a combination of cytokines including IL-1. The RNA was pooled for polyA RNA isolation and library construction.
EOSIHET02	PBLUESCRIPT	Library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
LIVRNON08	pINCY	This normalized liver tissue library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was isolated from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al. Proc. Natl. Acad. Sci. USA (1994) 91:9228 and Bonaldo et al. (1996) Genome Research 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.

Table 6 (cont.)

Library	Vector	Library Description
LUNGTUT06	pINCY	Library was constructed using RNA isolated from apical lung tumor tissue removed from an 80-year-old Caucasian female during a segmental lung resection. Pathology indicated a metastatic granulosa cell tumor. Patient history included pelvic soft tissue tumor and chemotherapy for one year. Family history included tuberculousis, lung cancer, and atherosclerotic coronary artery disease.
PROSNOT18	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrectomy. Pathology indicated adenofibromatous hyperplasia; this tissue was associated with a grade 3 transitional cell carcinoma. Patient history included angina and emphysema. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 1 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 1 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) Nucleic Acids Res. 19:1954 and Bonaldo et al. Genome Research (1996) 6:791.
SINTNOT18	pINCY	Library was constructed using RNA isolated from small intestine tissue obtained from a 59-year-old male.
COLENOR03	PCDNA2.1	Library was constructed using RNA isolated from colon epithelium tissue removed from a 13-year-old Caucasian female who died from a motor vehicle accident.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
PASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- 5           a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-27,  
            b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27,  
            c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and  
10           d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-27.

15           3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

20           5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:28-54.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

25           7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30           9. A method for producing a polypeptide of claim 1, the method comprising:

            a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

35           b) recovering the polypeptide so expressed.



10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5           a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54,  
            b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54,  
            c) a polynucleotide sequence complementary to a),  
            d) a polynucleotide sequence complementary to b), and  
10           e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15           13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization  
20           complex is formed between said probe and said target polynucleotide or fragments thereof, and  
            b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25           15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
30           b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

18. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a

compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- 5           a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in
- 10          the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

15          comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of
- 20          the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at
- 25          least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- 30          d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.  
 BAUGHN, Mariah R.  
 BURFORD, Neil  
 AU-YOUNG, Janice  
 LU, Dyung Aina M.  
 YANG, Junming  
 REDDY, Roopa  
 LAL, Preeti  
 HILLMAN, Jennifer L.  
 AZIMZAI, Yalda  
 YUE, Henry  
 NGUYEN, Danniell B.  
 YAO, Monique G.  
 GANDHI, Ameen R.  
 TANG, Y. Tom  
 KHAN, Farrah A.

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0005 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/172,000; 60/176,083; 60/177,332; 60/178,572; 60/179,758;  
 60/181,625

<151> 1999-12-23; 2000-01-14; 2000-01-21; 2000-01-28; 2000-02-02;  
 2000-02-10

<160> 54

<170> PERL Program

<210> 1

<211> 477

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1416107CD1

<400> 1

Met	Thr	Pro	Glu	Asp	Pro	Glu	Glu	Thr	Gln	Pro	Leu	Leu	Gly	Pro
1				5					10					15
Pro	Gly	Gly	Ser	Ala	Pro	Arg	Gly	Arg	Arg	Val	Phe	Leu	Ala	Ala
				20					25					30
Phe	Ala	Ala	Ala	Leu	Gly	Pro	Leu	Ser	Phe	Gly	Phe	Ala	Leu	Gly
				35					40					45
Tyr	Ser	Ser	Pro	Ala	Ile	Pro	Ser	Leu	Gln	Arg	Ala	Ala	Pro	Pro
				50					55					60
Ala	Pro	Arg	Leu	Asp	Asp	Ala	Ala	Ala	Ser	Trp	Phe	Gly	Ala	Val
				65					70					75
Val	Thr	Leu	Gly	Ala	Ala	Ala	Gly	Gly	Val	Leu	Gly	Gly	Trp	Leu
				80					85					90
Val	Asp	Arg	Ala	Gly	Arg	Lys	Leu	Ser	Leu	Leu	Leu	Cys	Ser	Val
				95					100					105
Pro	Phe	Val	Ala	Gly	Phe	Ala	Val	Ile	Thr	Ala	Ala	Gln	Asp	Val
				110					115					120
Trp	Met	Leu	Leu	Gly	Gly	Arg	Leu	Leu	Thr	Gly	Leu	Ala	Cys	Gly
				125					130					135
Val	Ala	Ser	Leu	Val	Ala	Pro	Val	Tyr	Ile	Ser	Glu	Ile	Ala	Tyr
				140					145					150
Pro	Ala	Val	Arg	Gly	Leu	Leu	Gly	Ser	Cys	Val	Gln	Leu	Met	Val
				155					160					165
Val	Val	Gly	Ile	Leu	Leu	Ala	Tyr	Leu	Ala	Gly	Trp	Val	Leu	Glu
				170					175					180
Trp	Arg	Trp	Leu	Ala	Val	Leu	Gly	Cys	Val	Pro	Pro	Ser	Leu	Met

Leu	Leu	Leu	Met	185	Cys	Phe	Met	Pro	Glu	190	Thr	Pro	Arg	Phe	Leu	195	Leu
Thr	Gln	His	Arg	200	Arg	Gln	Glu	Ala	Met	205	Ala	Ala	Leu	Arg	Phe	210	Leu
Trp	Gly	Ser	Glu	215	Gln	Gly	Trp	Glu	Asp	220	Pro	Pro	Ile	Gly	Ala	225	Glu
Gln	Ser	Phe	His	230	Leu	Ala	Leu	Leu	Arg	235	Gln	Pro	Gly	Ile	Tyr	240	Lys
Pro	Phe	Ile	Ile	245	Gly	Val	Ser	Leu	Met	250	Ala	Phe	Gln	Gln	Leu	255	Ser
Gly	Val	Asn	Ala	260	Val	Met	Phe	Tyr	Ala	265	Glu	Thr	Ile	Phe	Glu	270	Glu
Ala	Lys	Phe	Lys	275	Asp	Ser	Ser	Leu	Ala	280	Ser	Val	Val	Val	Gly	285	Val
Ile	Gln	Val	Leu	290	Phe	Thr	Ala	Val	Ala	295	Ala	Leu	Ile	Met	Asp	300	Arg
Ala	Gly	Arg	Arg	305	Leu	Leu	Leu	Val	Leu	310	Ser	Gly	Val	Val	Met	315	Val
Phe	Ser	Thr	Ser	320	Ala	Phe	Gly	Ala	Tyr	325	Phe	Lys	Leu	Thr	Gln	330	Gly
Gly	Pro	Gly	Asn	335	Ser	Ser	His	Val	Ala	340	Ile	Ser	Ala	Pro	Val	345	Ser
Ala	Gln	Pro	Val	350	Asp	Ala	Ser	Val	Gly	355	Leu	Ala	Trp	Leu	Ala	360	Val
Gly	Ser	Met	Cys	365	Leu	Phe	Ile	Ala	Gly	370	Phe	Ala	Val	Gly	Trp	375	Gly
Pro	Ile	Pro	Trp	380	Leu	Leu	Met	Ser	Glu	385	Ile	Phe	Pro	Leu	His	390	Val
Lys	Gly	Val	Ala	395	Thr	Gly	Ile	Cys	Val	400	Leu	Thr	Asn	Trp	Leu	405	Met
Ala	Phe	Leu	Val	410	Thr	Lys	Glu	Phe	Ser	415	Ser	Leu	Met	Glu	Val	420	Leu
Arg	Pro	Tyr	Gly	425	Ala	Phe	Trp	Leu	Ala	430	Ser	Ala	Phe	Cys	Ile	435	Phe
Ser	Val	Leu	Phe	440	Thr	Leu	Phe	Cys	Val	445	Pro	Glu	Thr	Lys	Gly	450	Lys
Thr	Leu	Glu	Gln	455	Ile	Thr	Ala	His	Phe	460	Glu	Gly	Arg			465	
				470						475							

&lt;210&gt; 2

&lt;211&gt; 498

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1682513CD1

&lt;400&gt; 2

Met	Arg	Arg	Gln	Asp	Ser	Arg	Gly	Asn	Thr	Val	Leu	His	Ala	Leu
1				5					10					15
Val	Ala	Ile	Ala	Asp	Asn	Thr	Arg	Glu	Asn	Thr	Lys	Phe	Val	Thr
				20					25					30
Lys	Met	Tyr	Asp	Leu	Leu	Leu	Leu	Lys	Cys	Ala	Arg	Leu	Phe	Pro
				35					40					45
Asp	Ser	Asn	Leu	Glu	Ala	Val	Leu	Asn	Asn	Asp	Gly	Leu	Ser	Pro
				50					55					60
Leu	Met	Met	Ala	Ala	Lys	Thr	Gly	Lys	Ile	Gly	Asn	Arg	His	Glu
				65					70					75
Met	Leu	Ala	Val	Glu	Pro	Ile	Asn	Glu	Leu	Leu	Arg	Asp	Lys	Trp
				80					85					90
Arg	Lys	Phe	Gly	Ala	Val	Ser	Phe	Tyr	Ile	Asn	Val	Val	Ser	Tyr
				95					100					105
Leu	Cys	Ala	Met	Val	Ile	Phe	Thr	Leu	Thr	Ala	Tyr	Tyr	Gln	Pro
				110					115					120
Leu	Glu	Gly	Thr	Pro	Pro	Tyr	Pro	Tyr	Arg	Thr	Thr	Val	Asp	Tyr

Leu Arg Leu Ala	125	Gly Glu Val Ile Thr	130	Leu Phe Thr Gly Val	135
Phe Phe Phe Thr	140	Asn Ile Lys Asp Leu	145	Met Lys Lys Cys	150
Gly Val Asn Ser	155	Leu Phe Ile Asp Gly	160	Ser Phe Gln Leu Leu	165
Phe Ile Tyr Ser	170	Val Leu Val Ile Val	175	Ser Ala Ala Leu Tyr	180
Ala Gly Ile Glu	185	Ala Tyr Leu Ala Val	190	Met Val Phe Ala Leu	195
Leu Gly Trp Met	200	Asn Ala Leu Tyr Phe	205	Thr Arg Gly Leu Lys	210
Thr Gly Thr Tyr	215	Ser Ile Met Ile Gln	220	Lys Ile Leu Phe Lys	225
Leu Phe Arg Phe	230	Leu Leu Val Tyr Leu	235	Leu Phe Met Ile Gly	240
Ala Ser Ala Leu	245	Val Ser Leu Leu Asn	250	Pro Cys Ala Asn Met	255
Val Cys Asn Gly	260	Asp Gln Thr Asn Cys	265	Thr Val Pro Thr Tyr	270
Ser Cys Arg Asp	275	Ser Glu Thr Phe Ser	280	Thr Phe Leu Leu Asp	285
Phe Lys Leu Thr	290	Ile Gly Met Gly Asp	295	Leu Glu Met Leu Ser	300
Thr Lys Tyr Pro	305	Val Val Phe Ile Ile	310	Leu Leu Val Thr Tyr	315
Ile Leu Thr Phe	320	Val Leu Leu Leu Asn	325	Met Leu Ile Ala Leu	330
Gly Glu Thr Val	335	Gly Gln Val Ser Lys	340	Glu Ser Lys His Ile	345
Lys Leu Gln Trp	350	Ala Thr Thr Ile Leu	355	Asp Ile Glu Arg Ser	360
Pro Val Phe Leu	365	Arg Lys Ser Phe Arg	370	Ser Gly Glu Met Val	375
Val Gly Lys Ser	380	Ser Asp Gly Thr Pro	385	Asp Arg Arg Trp Cys	390
Arg Val Asp Glu	395	Val Asn Trp Ser His	400	Trp Asn Gln Asn Leu	405
Ile Ile Asn Glu	410	Asp Pro Gly Lys Asn	415	Glu Thr Tyr Gln Tyr	420
Gly Phe Ser His	425	Thr Val Gly Arg Leu	430	Arg Arg Asp Arg Trp	435
Ser Val Val Pro	440	Arg Val Val Glu Leu	445	Asn Lys Asn Ser Asn	450
Asp Glu Val Val	455	Val Pro Leu Asp Ser	460	Thr Gly Asn Pro Arg	465
Asp Gly His Gln	470	Gln Gly Tyr Pro Arg	475	Lys Trp Arg Thr Asp	480
Ala Pro Leu	485		490		495

&lt;210&gt; 3

&lt;211&gt; 764

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2446438CD1

&lt;400&gt; 3

Met Thr Ser Pro	Ser Ser Ser Pro	Val Phe Arg Leu	Glu Thr Leu
1	5	10	15
Asp Ala Gly Gln	Glu Asp Gly Ser	Glu Ala Asp Arg	Gly Lys Leu
20	25	30	
Asp Phe Gly Ser	Gly Leu Pro Pro	Met Glu Ser	Gln Phe Gln Gly

				35					40					45
Glu	Asp	Arg	Lys	Phe	Ala	Pro	Gln	Ile	Arg	Val	Asn	Leu	Asn	Tyr
				50					55					60
Arg	Lys	Gly	Thr	Gly	Ala	Ser	Gln	Pro	Asp	Pro	Asn	Arg	Phe	Asp
				65					70					75
Arg	Asp	Arg	Leu	Phe	Asn	Ala	Val	Ser	Arg	Gly	Val	Pro	Glu	Asp
				80					85					90
Leu	Ala	Gly	Leu	Pro	Glu	Tyr	Leu	Ser	Lys	Thr	Ser	Lys	Tyr	Leu
				95					100					105
Thr	Asp	Ser	Glu	Tyr	Thr	Glu	Gly	Ser	Thr	Gly	Lys	Thr	Cys	Leu
				110					115					120
Met	Lys	Ala	Val	Leu	Asn	Leu	Lys	Asp	Gly	Val	Asn	Ala	Cys	Ile
				125					130					135
Leu	Pro	Leu	Leu	Gln	Ile	Asp	Arg	Asp	Ser	Gly	Asn	Pro	Gln	Pro
				140					145					150
Leu	Val	Asn	Ala	Gln	Cys	Thr	Asp	Asp	Tyr	Tyr	Arg	Gly	His	Ser
				155					160					165
Ala	Leu	His	Ile	Ala	Ile	Glu	Lys	Arg	Ser	Leu	Gln	Cys	Val	Lys
				170					175					180
Leu	Leu	Val	Glu	Asn	Gly	Ala	Asn	Val	His	Ala	Arg	Ala	Cys	Gly
				185					190					195
Arg	Phe	Phe	Gln	Lys	Gly	Gln	Gly	Thr	Cys	Phe	Tyr	Phe	Gly	Glu
				200					205					210
Leu	Pro	Leu	Ser	Leu	Ala	Ala	Cys	Thr	Lys	Gln	Trp	Asp	Val	Val
				215					220					225
Ser	Tyr	Leu	Leu	Glu	Asn	Pro	His	Gln	Pro	Ala	Ser	Leu	Gln	Ala
				230					235					240
Thr	Asp	Ser	Gln	Gly	Asn	Thr	Val	Leu	His	Ala	Leu	Val	Met	Ile
				245					250					255
Ser	Asp	Asn	Ser	Ala	Glu	Asn	Ile	Ala	Leu	Val	Thr	Ser	Met	Tyr
				260					265					270
Asp	Gly	Leu	Leu	Gln	Ala	Gly	Ala	Arg	Leu	Cys	Pro	Thr	Val	Gln
				275					280					285
Leu	Glu	Asp	Ile	Arg	Asn	Leu	Gln	Asp	Leu	Thr	Pro	Leu	Lys	Leu
				290					295					300
Ala	Ala	Lys	Glu	Gly	Lys	Ile	Glu	Ile	Phe	Arg	His	Ile	Leu	Gln
				305					310					315
Arg	Glu	Phe	Ser	Gly	Leu	Ser	His	Leu	Ser	Arg	Lys	Phe	Thr	Glu
				320					325					330
Trp	Cys	Tyr	Gly	Pro	Val	Arg	Val	Ser	Leu	Tyr	Asp	Leu	Ala	Ser
				335					340					345
Val	Asp	Ser	Cys	Glu	Glu	Asn	Ser	Val	Leu	Glu	Ile	Ile	Ala	Phe
				350					355					360
His	Cys	Lys	Ser	Pro	His	Arg	His	Arg	Met	Val	Val	Leu	Glu	Pro
				365					370					375
Leu	Asn	Lys	Leu	Leu	Gln	Ala	Lys	Trp	Asp	Leu	Leu	Ile	Pro	Lys
				380					385					390
Phe	Phe	Leu	Asn	Phe	Leu	Cys	Asn	Leu	Ile	Tyr	Met	Phe	Ile	Phe
				395					400					405
Thr	Ala	Val	Ala	Tyr	His	Gln	Pro	Thr	Leu	Lys	Lys	Gln	Ala	Ala
				410					415					420
Pro	His	Leu	Lys	Ala	Glu	Val	Gly	Asn	Ser	Met	Leu	Leu	Thr	Gly
				425					430					435
His	Ile	Leu	Ile	Leu	Leu	Gly	Gly	Ile	Tyr	Leu	Leu	Val	Gly	Gln
				440					445					450
Leu	Trp	Tyr	Phe	Trp	Arg	Arg	His	Val	Phe	Ile	Trp	Ile	Ser	Phe
				455					460					465
Ile	Asp	Ser	Tyr	Phe	Glu	Ile	Leu	Phe	Leu	Phe	Gln	Ala	Leu	Leu
				470					475					480
Thr	Val	Val	Ser	Gln	Val	Leu	Cys	Phe	Leu	Ala	Ile	Glu	Trp	Tyr
				485					490					495
Leu	Pro	Leu	Leu	Val	Ser	Ala	Leu	Val	Leu	Gly	Trp	Leu	Asn	Leu
				500					505					510
Leu	Tyr	Tyr	Thr	Arg	Gly	Phe	Gln	His	Thr	Gly	Ile	Tyr	Ser	Val
				515					520					525
Met	Ile	Gln	Lys	Val	Ile	Leu	Arg	Asp	Leu	Leu	Arg	Phe	Leu	Leu
				530					535					540

Ile	Tyr	Leu	Val	Phe	Leu	Phe	Gly	Phe	Ala	Val	Ala	Leu	Val	Ser
				545					550					555
Leu	Ser	Gln	Glu	Ala	Trp	Arg	Pro	Glu	Ala	Pro	Thr	Gly	Pro	Asn
				560					565					570
Ala	Thr	Glu	Ser	Val	Gln	Pro	Met	Glu	Gly	Gln	Glu	Asp	Glu	Gly
				575					580					585
Asn	Gly	Ala	Gln	Tyr	Arg	Gly	Ile	Leu	Glu	Ala	Ser	Leu	Glu	Leu
				590					595					600
Phe	Lys	Phe	Thr	Ile	Gly	Met	Gly	Glu	Leu	Ala	Phe	Gln	Glu	Gln
				605					610					615
Leu	His	Phe	Arg	Gly	Met	Val	Leu	Leu	Leu	Leu	Leu	Ala	Tyr	Val
				620					625					630
Leu	Leu	Thr	Tyr	Ile	Leu	Leu	Leu	Asn	Met	Leu	Ile	Ala	Leu	Met
				635					640					645
Ser	Glu	Thr	Val	Asn	Ser	Val	Ala	Thr	Asp	Ser	Trp	Ser	Ile	Trp
				650					655					660
Lys	Leu	Gln	Lys	Ala	Ile	Ser	Val	Leu	Glu	Met	Glu	Asn	Gly	Tyr
				665					670					675
Trp	Trp	Cys	Arg	Lys	Lys	Gln	Arg	Ala	Gly	Val	Met	Leu	Thr	Val
				680					685					690
Gly	Thr	Lys	Pro	Asp	Gly	Ser	Pro	Asp	Glu	Arg	Trp	Cys	Phe	Arg
				695					700					705
Val	Glu	Glu	Val	Asn	Trp	Ala	Ser	Trp	Glu	Gln	Thr	Leu	Pro	Thr
				710					715					720
Leu	Cys	Glu	Asp	Pro	Ser	Gly	Ala	Gly	Val	Pro	Arg	Thr	Leu	Glu
				725					730					735
Asn	Pro	Val	Leu	Ala	Ser	Pro	Pro	Lys	Glu	Asp	Glu	Asp	Gly	Ala
				740					745					750
Ser	Glu	Glu	Asn	Tyr	Val	Pro	Val	Gln	Leu	Leu	Gln	Ser	Asn	
				755					760					

&lt;210&gt; 4

&lt;211&gt; 255

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2817822CD1

&lt;400&gt; 4

Met	Trp	Gln	Gly	Cys	Ala	Val	Glu	Arg	Pro	Val	Gly	Arg	Met	Thr
1				5					10					15
Ser	Gln	Thr	Pro	Leu	Pro	Gln	Ser	Pro	Arg	Pro	Arg	Arg	Pro	Thr
				20					25					30
Met	Ser	Thr	Val	Val	Glu	Leu	Asn	Val	Gly	Gly	Glu	Phe	His	Thr
				35					40					45
Thr	Thr	Leu	Gly	Thr	Leu	Arg	Lys	Phe	Pro	Gly	Ser	Lys	Leu	Ala
				50					55					60
Glu	Met	Phe	Ser	Ser	Leu	Ala	Lys	Ala	Ser	Thr	Asp	Ala	Glu	Gly
				65					70					75
Arg	Phe	Phe	Ile	Asp	Arg	Pro	Ser	Thr	Tyr	Phe	Arg	Pro	Ile	Leu
				80					85					90
Asp	Tyr	Leu	Arg	Thr	Gly	Gln	Val	Pro	Thr	Gln	His	Ile	Pro	Glu
				95					100					105
Val	Tyr	Arg	Glu	Ala	Gln	Phe	Tyr	Glu	Ile	Lys	Pro	Leu	Val	Lys
				110					115					120
Leu	Leu	Glu	Asp	Met	Pro	Gln	Ile	Phe	Gly	Glu	Gln	Val	Ser	Arg
				125					130					135
Lys	Gln	Phe	Leu	Leu	Gln	Val	Pro	Gly	Tyr	Ser	Glu	Asn	Leu	Glu
				140					145					150
Leu	Met	Val	Arg	Leu	Ala	Arg	Ala	Glu	Ala	Ile	Thr	Ala	Arg	Lys
				155					160					165
Ser	Ser	Val	Leu	Val	Cys	Leu	Val	Glu	Thr	Glu	Glu	Gln	Asp	Ala
				170					175					180
Tyr	Tyr	Ser	Glu	Val	Leu	Cys	Phe	Leu	Gln	Asp	Lys	Lys	Met	Phe
				185					190					195



Lys	Ser	Val	Val	Lys	Phe	Gly	Pro	Trp	Lys	Ala	Val	Leu	Asp	Asn
				200					205					210
Ser	Asp	Leu	Met	His	Cys	Leu	Glu	Met	Asp	Ile	Lys	Ala	Gln	Gly
				215					220					225
Tyr	Lys	Val	Phe	Ser	Lys	Phe	Tyr	Leu	Thr	Tyr	Pro	Thr	Lys	Arg
				230					235					240
Asn	Glu	Phe	His	Phe	Asn	Ile	Tyr	Ser	Phe	Thr	Phe	Thr	Trp	Trp
				245					250					255

&lt;210&gt; 5

&lt;211&gt; 584

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4009329CD1

&lt;400&gt; 5

Met	Ala	Gly	Arg	Arg	Leu	Asn	Leu	Arg	Trp	Ala	Leu	Ser	Val	Leu
1				5					10					15
Cys	Val	Leu	Leu	Met	Ala	Glu	Thr	Val	Ser	Gly	Thr	Arg	Gly	Ser
				20					25					30
Ser	Thr	Gly	Ala	His	Ile	Ser	Pro	Gln	Phe	Pro	Ala	Ser	Gly	Val
				35					40					45
Asn	Gln	Thr	Pro	Val	Val	Asp	Cys	Arg	Lys	Val	Cys	Gly	Leu	Asn
				50					55					60
Val	Ser	Asp	Arg	Cys	Asp	Phe	Ile	Arg	Thr	Asn	Pro	Asp	Cys	His
				65					70					75
Ser	Asp	Gly	Gly	Tyr	Leu	Asp	Tyr	Leu	Glu	Gly	Ile	Phe	Cys	His
				80					85					90
Phe	Pro	Pro	Ser	Leu	Leu	Pro	Leu	Ala	Val	Thr	Leu	Tyr	Val	Ser
				95					100					105
Trp	Leu	Leu	Tyr	Leu	Phe	Leu	Ile	Leu	Gly	Val	Thr	Ala	Ala	Lys
				110					115					120
Phe	Phe	Cys	Pro	Asn	Leu	Ser	Ala	Ile	Ser	Thr	Thr	Leu	Lys	Leu
				125					130					135
Ser	His	Asn	Val	Ala	Gly	Val	Thr	Phe	Leu	Ala	Phe	Gly	Asn	Gly
				140					145					150
Ala	Pro	Asp	Ile	Phe	Ser	Ala	Leu	Val	Ala	Phe	Ser	Asp	Pro	His
				155					160					165
Thr	Ala	Gly	Leu	Ala	Leu	Gly	Ala	Leu	Phe	Gly	Ala	Gly	Val	Leu
				170					175					180
Val	Thr	Thr	Val	Val	Ala	Gly	Gly	Ile	Thr	Ile	Leu	His	Pro	Phe
				185					190					195
Met	Ala	Ala	Ser	Arg	Pro	Phe	Phe	Arg	Asp	Ile	Val	Phe	Tyr	Met
				200					205					210
Val	Ala	Val	Phe	Leu	Thr	Phe	Leu	Met	Leu	Phe	Arg	Gly	Arg	Val
				215					220					225
Thr	Leu	Ala	Trp	Ala	Leu	Gly	Tyr	Leu	Gly	Leu	Tyr	Val	Phe	Tyr
				230					235					240
Val	Val	Thr	Val	Ile	Leu	Cys	Thr	Trp	Ile	Tyr	Gln	Arg	Gln	Arg
				245					250					255
Arg	Gly	Ser	Leu	Phe	Cys	Pro	Met	Pro	Val	Thr	Pro	Glu	Ile	Leu
				260					265					270
Ser	Asp	Ser	Glu	Glu	Asp	Arg	Val	Ser	Ser	Asn	Thr	Asn	Ser	Tyr
				275					280					285
Asp	Tyr	Gly	Asp	Glu	Tyr	Arg	Pro	Leu	Phe	Phe	Tyr	Gln	Glu	Thr
				290					295					300
Thr	Ala	Gln	Ile	Leu	Val	Arg	Ala	Leu	Asn	Pro	Leu	Asp	Tyr	Met
				305					310					315
Lys	Trp	Arg	Arg	Lys	Ser	Ala	Tyr	Trp	Lys	Ala	Leu	Lys	Val	Phe
				320					325					330
Lys	Leu	Pro	Val	Glu	Phe	Leu	Leu	Leu	Leu	Thr	Val	Pro	Val	Val
				335					340					345
Asp	Pro	Asp	Lys	Asp	Asp	Gln	Asn	Trp	Lys	Arg	Pro	Leu	Asn	Cys

Leu His Leu Val	350	Ile Ser Pro Leu Val	355	Val Val Leu Thr Leu	360
Ser Gly Thr Tyr	365	Gly Val Tyr Glu Ile	370	Gly Gly Leu Val Pro	375
Trp Val Val Val	380	Val Ile Ala Gly Thr	385	Ala Leu Ala Ser Val	390
Phe Phe Ala Thr	395	Ser Asp Ser Gln Pro	400	Pro Arg Leu His Trp	405
Phe Ala Phe Leu	410	Gly Phe Leu Thr Ser	415	Ala Leu Trp Ile Asn	420
Ala Ala Thr Glu	425	Val Val Asn Ile Leu	430	Arg Ser Leu Gly Val	435
Phe Arg Leu Ser	440	Asn Thr Val Leu Gly	445	Leu Thr Leu Leu Ala	450
Gly Asn Ser Ile	455	Gly Asp Ala Phe Ser	460	Asp Phe Thr Leu Ala	465
Gln Gly Tyr Pro	470	Arg Met Ala Phe Ser	475	Ala Cys Phe Gly Gly	480
Ile Phe Asn Ile	485	Leu Val Gly Val Gly	490	Leu Gly Cys Leu Leu	495
Ile Ser Arg Ser	500	His Thr Glu Val Lys	505	Leu Glu Pro Asp Gly	510
Leu Val Trp Val	515	Leu Ala Gly Ala Leu	520	Gly Leu Ser Leu Val	525
Ser Leu Val Ser	530	Val Pro Leu Gln Cys	535	Phe Gln Leu Ser Arg	540
Tyr Gly Phe Cys	545	Leu Leu Phe Tyr	550	Leu Asn Phe Leu Val	555
Ala Leu Leu Ile	560	Glu Phe Gly Val Ile	565	His Leu Lys Ser Met	570
	575		580		

&lt;210&gt; 6

&lt;211&gt; 416

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6618083CD1

&lt;400&gt; 6

Met Lys Leu Ser Lys	5	Lys Asp Arg Gly Glu	10	Asp Glu Glu Ser Asp	15
Ser Ala Lys Lys Lys	20	Leu Asp Trp Ser Cys	25	Ser Leu Leu Val Ala	30
Ser Leu Ala Gly Ala	35	Phe Gly Ser Ser Phe	40	Leu Tyr Gly Tyr Asn	45
Leu Ser Val Val Asn	50	Ala Pro Thr Pro Tyr	55	Ile Lys Ala Phe Tyr	60
Asn Glu Ser Trp Glu	65	Arg Arg His Gly Arg	70	Pro Ile Asp Pro Asp	75
Thr Leu Thr Leu Leu	80	Trp Ser Val Thr Val	85	Ser Ile Phe Ala Ile	90
Gly Gly Leu Val Gly	95	Thr Leu Ile Val Lys	100	Met Ile Gly Lys Val	105
Leu Gly Arg Lys His	110	Thr Leu Leu Ala Asn	115	Asn Gly Phe Ala Ile	120
Ser Ala Ala Leu Leu	125	Met Ala Cys Ser Leu	130	Gln Ala Gly Ala Phe	135
Glu Met Leu Ile Val	140	Gly Arg Phe Ile Met	145	Gly Ile Asp Gly Gly	150
Val Ala Leu Ser Val	155	Leu Pro Met Tyr Leu	160	Ser Glu Ile Ser Pro	165
Lys Glu Ile Arg Gly	170	Ser Leu Gly Gln Val	175	Thr Ala Ile Phe Ile	180
Cys Ile Gly Val Phe		Thr Gly Gln Leu Leu		Gly Leu Pro Glu Leu	

Leu Gly Lys Glu	185	Thr Trp Pro Tyr	190	Phe Gly Val Ile	195
Val Pro Ala Val	200	Val Gln Leu Leu Ser	205	Pro Phe Leu Pro	210
Ser Pro Arg Tyr	215	Leu Leu Leu Glu Lys	220	His Asn Glu Ala Arg	225
Val Lys Ala Phe	230	Gln Thr Phe Leu Gly	235	Ala Asp Val Ser	240
Glu Val Glu Glu	245	Val Leu Ala Glu Ser	250	Val Gln Arg Ser	255
Arg Leu Val Ser	260	Val Leu Glu Leu Leu	265	Ala Pro Tyr Val	270
Trp Gln Val Val	275	Thr Val Ile Val Thr	280	Met Ala Cys Tyr Gln	285
Cys Gly Leu Asn	290	Ala Ile Trp Phe Tyr	295	Thr Asn Ser Ile Phe	300
Lys Ala Gly Ile	305	Pro Leu Ala Lys Ile	310	Tyr Val Thr Leu Ser	315
Thr Gly Gly Ile	320	Glu Thr Leu Ala Ala	325	Val Phe Ser Gly Leu	330
Ile Glu His Leu	335	Gly Arg Arg Pro Leu	340	Ile Gly Gly Phe	345
Leu Met Gly Leu	350	Phe Phe Gly Thr Leu	355	Thr Ile Thr Leu Thr	360
Gln Asp His Ala	365	Pro Trp Val Pro Tyr	370	Leu Ser Ile Val Gly	375
Leu Ala Ile Ile	380	Ala Ser Phe Cys Ser	385	Pro Ala Val Phe	390
Glu Glu Thr Val	395	Asn Val Ser Ile Val	400	Ser Glu	405
	410		415		

&lt;210&gt; 7

&lt;211&gt; 664

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472002CD1

&lt;400&gt; 7

Met Thr Glu Lys Thr	Asn Gly Val Lys Ser	Ser Pro Ala Asn Asn
1	5	10
His Asn His His Ala	Pro Pro Ala Ile Lys	Ala Asn Gly Lys Asp
20	25	30
Asp His Arg Thr Ser	Ser Arg Pro His Ser	Ala Ala Asp Asp Asp
35	40	45
Thr Ser Ser Glu Leu	Gln Arg Leu Ala Asp	Val Asp Ala Pro Gln
50	55	60
Gln Gly Arg Ser Gly	Phe Arg Arg Ile Val	Arg Leu Val Gly Ile
65	70	75
Ile Arg Glu Trp Ala	Asn Lys Asn Phe Arg	Glu Glu Glu Pro Arg
80	85	90
Pro Asp Ser Phe Leu	Glu Arg Phe Arg Gly	Pro Glu Leu Gln Thr
95	100	105
Val Thr Thr Gln Glu	Gly Asp Gly Lys Gly	Asp Lys Asp Gly Glu
110	115	120
Asp Lys Gly Thr Lys	Lys Lys Phe Glu Leu	Phe Val Leu Asp Pro
125	130	135
Ala Gly Asp Trp Tyr	Tyr Cys Trp Leu Phe	Val Ile Ala Met Pro
140	145	150
Val Leu Tyr Asn Trp	Cys Leu Leu Val Ala	Arg Ala Cys Phe Ser
155	160	165
Asp Leu Gln Lys Gly	Tyr Tyr Leu Val Trp	Leu Val Leu Asp Tyr
170	175	180
Val Ser Asp Val Val	Tyr Ile Ala Asp Leu	Phe Ile Arg Leu Arg

Thr Gly Phe Leu	185	Gln Gly Leu Leu	190	Lys Asp Thr Lys	195
Leu Arg Asp Asn	200	Ile His Thr Leu	205	Phe Lys Leu Asp	210
Ala Ser Ile Ile	215	Pro Thr Asp Leu	220	Phe Ala Val Asp	225
His Ser Pro Glu	230	Val Arg Phe Asn	235	Leu His Phe Ala	240
Met Phe Glu Phe	245	Phe Asp Arg Thr	250	Thr Arg Thr Asn	255
Asn Ile Phe Arg	260	Ile Ser Asn Leu	265	Leu Tyr Ile Leu	270
Ile His Trp Asn	275	Ala Cys Ile Tyr	280	Ile Ser Lys Ser	285
Gly Phe Gly Val	290	Asp Thr Trp Val	295	Pro Asn Ile Thr	300
Glu Tyr Gly Tyr	305	Leu Ala Arg Glu	310	Tyr Cys Leu Tyr	315
Ser Thr Leu Thr	320	Leu Thr Thr Ile	325	Glu Thr Pro Pro	330
Lys Asp Glu Glu	335	Tyr Leu Phe Val	340	Phe Asp Phe Leu	345
Val Leu Ile Phe	350	Ala Thr Ile Val	355	Asn Val Gly Ser	360
Ser Asn Met Asn	365	Ala Thr Arg Ala	370	Phe Gln Ala Lys	375
Ala Val Lys His	380	Tyr Met Gln Phe	385	Val Ser Lys Gly	390
Glu Ala Lys Val	395	Ile Arg Trp Phe	400	Tyr Leu Trp Thr	405
Lys Thr Val Asp	410	Glu Arg Glu Ile	415	Leu Trp Thr Asn	420
Leu Arg Ala Glu	425	Ile Ala Ile Asn	430	Lys Asn Leu Pro	435
Lys Val Arg Ile	440	Phe His Asp Cys	445	Leu Ser Thr Leu	450
Leu Val Leu Lys	455	Leu Arg Pro Gln	460	Ala Gly Leu Leu	465
Ile Cys Arg Lys	470	Gly Asp Ile Gly	475	Phe Ser Pro Gly	480
Glu Gly Lys Leu	485	Ala Val Val Ala	490	Glu Met Tyr Ile	495
Ala Leu Leu Ser	500	Ala Gly Ser Cys	505	Asp Gly Val Thr	510
Asn Ile Lys Gly	515	Ser Lys Met Gly	520	Glu Ile Ser Ile	525
Arg Ser Leu Gly	530	Tyr Ser Asp Leu	535	Arg Arg Thr Ala	540
Leu Met Glu Ala	545	Val Thr Glu Tyr	550	Cys Leu Ser Lys	555
Glu Glu Arg Gly	560	Arg Glu Ile Leu	565	Asp Ala Lys Lys	570
Glu Asn Glu Val	575	Ala Thr Ser Met	580	Glu Gly Leu Leu	585
Leu Gly Gln Leu	590	Glu Thr Asn Met	595	Val Asp Val Gln	600
Gly Arg Leu Leu	605	Ala Glu Tyr Thr	610	Leu Tyr Thr Arg	615
Gln Arg Ile Thr	620	Val Leu Glu Thr	625	Ala Gln Gln Lys	630
Asp Asp Tyr Leu	635	Ser Asp Gly Met	640	Met Lys Gln Asn	645
Ala Asp Glu Pro	650	Asn Ser Pro Glu	655	Leu Ala Ala	660

&lt;210&gt; 8

<211> 242  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1812692CD1

<400> 8

Met	Ser	Phe	Arg	Ala	Ala	Arg	Leu	Ser	Met	Arg	Asn	Arg	Arg	Asn
1				5					10					15
Asp	Thr	Leu	Asp	Ser	Thr	Arg	Thr	Leu	Tyr	Ser	Ser	Ala	Ser	Arg
				20					25					30
Ser	Thr	Asp	Leu	Ser	Tyr	Ser	Glu	Ser	Asp	Leu	Val	Asn	Phe	Ile
				35					40					45
Gln	Ala	Asn	Phe	Lys	Lys	Arg	Glu	Cys	Val	Phe	Phe	Thr	Lys	Asp
				50					55					60
Ser	Lys	Ala	Thr	Glu	Asn	Val	Cys	Lys	Cys	Gly	Tyr	Ala	Gln	Ser
				65					70					75
Gln	His	Met	Glu	Gly	Thr	Gln	Ile	Asn	Gln	Ser	Glu	Lys	Trp	Asn
				80					85					90
Tyr	Lys	Lys	His	Thr	Lys	Glu	Phe	Pro	Thr	Asp	Ala	Phe	Gly	Asp
				95					100					105
Ile	Gln	Phe	Glu	Thr	Leu	Gly	Lys	Lys	Gly	Lys	Tyr	Ile	Arg	Leu
				110					115					120
Ser	Cys	Asp	Thr	Asp	Ala	Glu	Ile	Leu	Tyr	Glu	Leu	Leu	Thr	Gln
				125					130					135
His	Trp	His	Leu	Lys	Thr	Pro	Asn	Leu	Val	Ile	Ser	Val	Thr	Gly
				140					145					150
Gly	Ala	Lys	Asn	Phe	Ala	Leu	Lys	Pro	Arg	Met	Arg	Lys	Ile	Phe
				155					160					165
Ser	Arg	Leu	Ile	Tyr	Ile	Ala	Gln	Ser	Lys	Gly	Ala	Trp	Ile	Leu
				170					175					180
Thr	Gly	Gly	Thr	His	Tyr	Gly	Leu	Met	Lys	Tyr	Ile	Gly	Glu	Val
				185					190					195
Val	Arg	Asp	Asn	Thr	Ile	Ser	Arg	Ser	Ser	Glu	Glu	Asn	Ile	Val
				200					205					210
Ala	Ile	Gly	Ile	Ala	Ala	Trp	Gly	Met	Val	Ser	Asn	Arg	Asp	Thr
				215					220					225
Leu	Ile	Arg	Asn	Cys	Asp	Ala	Glu	Val	Pro	Val	Gly	Gln	Glu	Glu
				230					235					240
Val	Cys													

<210> 9  
 <211> 398  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3232992CD1

<400> 9

Met	Val	Ala	Ala	Pro	Ile	Phe	Gly	Tyr	Leu	Gly	Asp	Arg	Phe	Asn
1				5					10					15
Arg	Lys	Val	Ile	Leu	Ser	Cys	Gly	Ile	Phe	Phe	Trp	Ser	Ala	Val
				20					25					30
Thr	Phe	Ser	Ser	Ser	Phe	Ile	Pro	Gln	Gln	Tyr	Phe	Trp	Leu	Leu
				35					40					45
Val	Leu	Ser	Arg	Gly	Leu	Val	Gly	Ile	Gly	Glu	Ala	Ser	Tyr	Ser
				50					55					60
Thr	Ile	Ala	Pro	Thr	Ile	Ile	Gly	Asp	Leu	Phe	Thr	Lys	Asn	Thr
				65					70					75
Arg	Thr	Leu	Met	Leu	Ser	Val	Phe	Tyr	Phe	Ala	Ile	Pro	Leu	Gly
				80					85					90
Ser	Gly	Leu	Gly	Tyr	Ile	Thr	Gly	Ser	Ser	Val	Lys	Gln	Ala	Ala

				95					100					105
Gly	Asp	Trp	His	Trp	Ala	Leu	Arg	Val	Ser	Pro	Val	Leu	Gly	Met
				110					115					120
Ile	Thr	Gly	Thr	Leu	Ile	Leu	Ile	Leu	Val	Pro	Ala	Thr	Lys	Arg
				125					130					135
Gly	His	Ala	Asp	Gln	Leu	Gly	Asp	Gln	Leu	Lys	Ala	Arg	Thr	Ser
				140					145					150
Trp	Leu	Arg	Asp	Met	Lys	Ala	Leu	Ile	Arg	Asn	Arg	Ser	Tyr	Val
				155					160					165
Phe	Ser	Ser	Leu	Ala	Thr	Ser	Ala	Val	Ser	Phe	Ala	Thr	Gly	Ala
				170					175					180
Leu	Gly	Met	Trp	Ile	Pro	Leu	Tyr	Leu	His	Arg	Ala	Gln	Val	Val
				185					190					195
Gln	Lys	Thr	Ala	Glu	Thr	Cys	Asn	Ser	Pro	Pro	Cys	Gly	Ala	Lys
				200					205					210
Asp	Ser	Leu	Ile	Phe	Gly	Ala	Ile	Thr	Cys	Phe	Thr	Gly	Phe	Leu
				215					220					225
Gly	Val	Val	Thr	Gly	Ala	Gly	Ala	Thr	Arg	Trp	Cys	Arg	Leu	Lys
				230					235					240
Thr	Gln	Arg	Ala	Asp	Pro	Leu	Val	Cys	Ala	Val	Gly	Met	Leu	Gly
				245					250					255
Ser	Ala	Ile	Phe	Ile	Cys	Leu	Ile	Phe	Val	Ala	Ala	Lys	Ser	Ser
				260					265					270
Ile	Val	Gly	Ala	Tyr	Ile	Cys	Ile	Phe	Val	Gly	Glu	Thr	Leu	Leu
				275					280					285
Phe	Ser	Asn	Trp	Ala	Ile	Thr	Ala	Asp	Ile	Leu	Met	Tyr	Val	Val
				290					295					300
Ile	Pro	Thr	Arg	Arg	Ala	Thr	Ala	Val	Ala	Leu	Gln	Ser	Phe	Thr
				305					310					315
Ser	His	Leu	Leu	Gly	Asp	Ala	Gly	Ser	Pro	Tyr	Leu	Ile	Gly	Phe
				320					325					330
Ile	Ser	Asp	Leu	Ile	Arg	Gln	Ser	Thr	Lys	Asp	Ser	Pro	Leu	Trp
				335					340					345
Glu	Phe	Leu	Ser	Leu	Gly	Tyr	Ala	Leu	Met	Leu	Cys	Pro	Phe	Val
				350					355					360
Val	Val	Leu	Gly	Gly	Met	Phe	Phe	Leu	Ala	Thr	Ala	Leu	Phe	Phe
				365					370					375
Val	Ser	Asp	Arg	Ala	Arg	Ala	Glu	Gln	Gln	Val	Asn	Gln	Leu	Ala
				380					385					390
Met	Pro	Pro	Ala	Ser	Val	Lys	Val							
				395										

&lt;210&gt; 10

&lt;211&gt; 553

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3358383CD1

&lt;400&gt; 10

Met	Ala	Phe	Gln	Asp	Leu	Leu	Gly	His	Ala	Gly	Asp	Leu	Trp	Arg
1				5					10					15
Phe	Gln	Ile	Leu	Gln	Thr	Val	Phe	Leu	Ser	Ile	Phe	Ala	Val	Ala
				20					25					30
Thr	Tyr	Leu	His	Phe	Met	Leu	Glu	Asn	Phe	Thr	Ala	Phe	Ile	Pro
				35					40					45
Gly	His	Arg	Cys	Trp	Val	His	Ile	Leu	Asp	Asn	Asp	Thr	Val	Ser
				50					55					60
Asp	Asn	Asp	Thr	Gly	Ala	Leu	Ser	Gln	Asp	Ala	Leu	Leu	Arg	Ile
				65					70					75
Ser	Ile	Pro	Leu	Asp	Ser	Asn	Met	Arg	Pro	Glu	Lys	Cys	Arg	Arg
				80					85					90
Phe	Val	His	Pro	Gln	Trp	Gln	Leu	Leu	His	Leu	Asn	Gly	Thr	Phe
				95					100					105
Pro	Asn	Thr	Ser	Asp	Ala	Asp	Met	Glu	Pro	Cys	Val	Asp	Gly	Trp

Val Tyr Asp Arg	110	Ile Ser Phe Ser Ser	115	Thr Ile Val Thr Glu	120
Asp Leu Val Cys	125	Asp Ser Gln Ser Leu	130	Ser Val Ala Lys	135
Val Phe Met Ala	140	Gly Met Met Val Gly	145	Ile Leu Gly Gly	150
Leu Ser Asp Arg	155	Phe Gly Arg Arg Phe	160	Leu Arg Trp Cys	165
Leu Gln Val Ala	170	Ile Val Gly Thr Cys	175	Ala Ala Leu Ala Pro	180
Phe Leu Ile Tyr	185	Cys Ser Leu Arg Phe	190	Ser Gly Ile Ala	195
Met Ser Leu Ile	200	Thr Asn Thr Ile Met	205	Ile Ala Glu Trp	210
Thr His Arg Phe	215	Gln Ala Met Gly Ile	220	Leu Gly Met Cys	225
Ser Gly Ile Ala	230	Phe Met Thr Leu Ala	235	Leu Ala Phe Ala	240
Arg Asp Trp His	245	Ile Leu Gln Leu Val	250	Val Ser Val Pro Tyr	255
Val Ile Phe Leu	260	Thr Ser Ser Trp Leu	265	Leu Glu Ser Ala Arg	270
Leu Ile Ile Asn	275	Asn Lys Pro Glu Glu	280	Gly Leu Lys Glu Leu	285
Lys Ala Ala His	290	Arg Ser Gly Met Lys	295	Asn Ala Arg Asp Thr	300
Thr Leu Glu Ile	305	Leu Lys Ser Thr Met	310	Lys Glu Leu Glu	315
Ala Gln Lys Lys	320	Lys Pro Ser Leu Cys	325	Glu Met Leu His Met	330
Asn Ile Cys Lys	335	Arg Ile Ser Leu Leu	340	Phe Thr Arg Phe	345
Asn Phe Met Ala	350	Tyr Phe Gly Leu Asn	355	Leu His Val Gln His	360
Gly Asn Asn Val	365	Phe Leu Leu Gln Thr	370	Leu Phe Gly Ala Val	375
Leu Leu Ala Asn	380	Cys Val Ala Pro Trp	385	Ala Leu Lys Tyr Met	390
Arg Arg Ala Ser	395	Gln Met Arg Leu Met	400	Tyr Leu Leu Ala Ile	405
Phe Met Ala Ile	410	Ile Phe Val Pro Gln	415	Glu Met Gln Thr Leu	420
Glu Val Leu Ala	425	Thr Leu Gly Leu Gly	430	Ala Ser Ala Leu Thr	435
Thr Leu Ala Phe	440	Ala His Gly Asn Glu	445	Val Ile Pro Thr Ile	450
Arg Ala Arg Ala	455	Met Gly Ile Asn Ala	460	Thr Phe Ala Asn Ile	465
Gly Ala Leu Ala	470	Pro Leu Met Met Ile	475	Leu Ser Val Tyr Ser	480
Pro Leu Pro Trp	485	Ile Ile Tyr Gly Val	490	Phe Pro Phe Ile Ser	495
Phe Ala Phe Leu	500	Leu Leu Pro Glu Thr	505	Arg Asn Lys Pro Leu	510
Asp Thr Ile Gln	515	Asp Glu Lys Asn Glu	520	Lys Asp Pro Arg	525
Pro Lys Gln Glu	530	Asp Pro Arg Val Glu	535	Val Thr Gln Phe	540
	545		550		

<210> 11  
 <211> 213  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 4250091CD1

&lt;400&gt; 11

Met	Ser	Ser	Gln	Glu	Leu	Val	Thr	Leu	Asn	Val	Gly	Gly	Lys	Ile
1				5					10					15
Phe	Thr	Thr	Arg	Phe	Ser	Thr	Ile	Lys	Gln	Phe	Pro	Ala	Ser	Arg
				20					25					30
Leu	Ala	Arg	Met	Leu	Asp	Gly	Arg	Asp	Gln	Glu	Phe	Lys	Met	Val
				35					40					45
Gly	Gly	Gln	Ile	Phe	Val	Asp	Arg	Asp	Gly	Asp	Leu	Phe	Ser	Phe
				50					55					60
Ile	Leu	Asp	Phe	Leu	Arg	Thr	His	Gln	Leu	Leu	Leu	Pro	Thr	Glu
				65					70					75
Phe	Ser	Asp	Tyr	Leu	Arg	Leu	Gln	Arg	Glu	Ala	Leu	Phe	Tyr	Glu
				80					85					90
Leu	Arg	Ser	Leu	Val	Asp	Leu	Leu	Asn	Pro	Tyr	Leu	Leu	Gln	Pro
				95					100					105
Arg	Pro	Ala	Leu	Val	Glu	Val	His	Phe	Leu	Ser	Arg	Asn	Thr	Gln
				110					115					120
Ala	Phe	Phe	Arg	Val	Phe	Gly	Ser	Cys	Ser	Lys	Thr	Ile	Glu	Met
				125					130					135
Leu	Thr	Gly	Arg	Ile	Thr	Val	Phe	Thr	Glu	Gln	Pro	Ser	Ala	Pro
				140					145					150
Thr	Trp	Asn	Gly	Asn	Phe	Phe	Pro	Pro	Gln	Met	Thr	Leu	Leu	Pro
				155					160					165
Leu	Pro	Pro	Gln	Arg	Pro	Ser	Tyr	His	Asp	Leu	Val	Phe	Gln	Cys
				170					175					180
Gly	Ser	Asp	Ser	Thr	Thr	Asp	Asn	Gln	Thr	Gly	Val	Arg	Tyr	Phe
				185					190					195
Val	Leu	Cys	Ser	Ile	Ser	Leu	Val	Tyr	Gln	Phe	Val	Met	Phe	Ser
				200					205					210
Leu	Lys	Thr												

&lt;210&gt; 12

&lt;211&gt; 476

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70064803CD1

&lt;400&gt; 12

Met	Ala	Gly	Ser	Asp	Thr	Ala	Pro	Phe	Leu	Ser	Gln	Ala	Asp	Asp
1				5					10					15
Pro	Asp	Asp	Gly	Pro	Val	Pro	Gly	Thr	Pro	Gly	Leu	Pro	Gly	Ser
				20					25					30
Thr	Gly	Asn	Pro	Lys	Ser	Glu	Glu	Pro	Glu	Val	Pro	Asp	Gln	Glu
				35					40					45
Gly	Leu	Gln	Arg	Ile	Thr	Gly	Leu	Ser	Pro	Gly	Arg	Ser	Ala	Leu
				50					55					60
Ile	Val	Ala	Val	Leu	Cys	Tyr	Ile	Asn	Leu	Leu	Asn	Tyr	Met	Asp
				65					70					75
Arg	Phe	Thr	Val	Ala	Gly	Val	Leu	Pro	Asp	Ile	Glu	Gln	Phe	Phe
				80					85					90
Asn	Ile	Gly	Asp	Ser	Ser	Ser	Gly	Leu	Ile	Gln	Thr	Val	Phe	Ile
				95					100					105
Ser	Ser	Tyr	Met	Val	Leu	Ala	Pro	Val	Phe	Gly	Tyr	Leu	Gly	Asp
				110					115					120
Arg	Tyr	Asn	Arg	Lys	Tyr	Leu	Met	Cys	Gly	Gly	Ile	Ala	Phe	Trp
				125					130					135
Ser	Leu	Val	Thr	Leu	Gly	Ser	Ser	Phe	Ile	Pro	Gly	Glu	His	Phe
				140					145					150
Trp	Leu	Leu	Leu	Leu	Thr	Arg	Gly	Leu	Val	Gly	Val	Gly	Glu	Ala
				155					160					165
Ser	Tyr	Ser	Thr	Ile	Ala	Pro	Thr	Leu	Ile	Ala	Asp	Leu	Phe	Val



170	175	180
Ala Asp Gln Arg Ser Arg Met Leu Ser	Ile Phe Tyr Phe Ala Ile	
185	190	195
Pro Val Gly Ser Gly Leu Gly Tyr Ile	Ala Gly Ser Lys Val Lys	
200	205	210
Asp Met Ala Gly Asp Trp His Trp Ala	Leu Arg Val Thr Pro Gly	
215	220	225
Leu Gly Val Val Ala Val Leu Leu Leu	Phe Leu Val Val Arg Glu	
230	235	240
Pro Pro Arg Gly Ala Val Glu Arg His	Ser Asp Leu Pro Pro Leu	
245	250	255
Asn Pro Thr Ser Trp Trp Ala Asp Leu	Arg Ala Leu Ala Arg Asn	
260	265	270
Leu Ile Phe Gly Leu Ile Thr Cys Leu	Thr Gly Val Leu Gly Val	
275	280	285
Gly Leu Gly Val Glu Ile Ser Arg Arg	Leu Arg His Ser Asn Pro	
290	295	300
Arg Ala Asp Pro Leu Val Cys Ala Thr	Gly Leu Leu Gly Ser Ala	
305	310	315
Pro Phe Leu Phe Leu Ser Leu Ala Cys	Ala Arg Gly Ser Ile Val	
320	325	330
Ala Thr Tyr Ile Phe Ile Phe Ile Gly	Glu Thr Leu Leu Ser Met	
335	340	345
Asn Trp Ala Ile Val Ala Asp Ile Leu	Leu Tyr Val Val Ile Pro	
350	355	360
Thr Arg Arg Ser Thr Ala Glu Ala Phe	Gln Ile Val Leu Ser His	
365	370	375
Leu Leu Gly Asp Ala Gly Ser Pro Tyr	Leu Ile Gly Leu Ile Ser	
380	385	390
Asp Arg Leu Arg Arg Asn Trp Pro Pro	Ser Phe Leu Ser Glu Phe	
395	400	405
Arg Ala Leu Gln Phe Ser Leu Met Leu	Cys Ala Phe Val Gly Ala	
410	415	420
Leu Gly Gly Ala Ala Phe Leu Gly Thr	Ala Ile Phe Ile Glu Ala	
425	430	435
Asp Arg Arg Arg Ala Gln Leu His Val	Gln Gly Leu Leu His Glu	
440	445	450
Ala Gly Ser Thr Asp Asp Arg Ile Val	Val Pro Gln Arg Gly Arg	
455	460	465
Ser Thr Arg Val Pro Val Ala Ser Val	Leu Ile	
470	475	

<210> 13  
 <211> 246  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 70356768CD1

<400> 13  
 Met Leu His Ala Leu Leu Arg Ser Arg Met Ile Gln Gly Arg Ile  
 1 5 10 15  
 Leu Leu Leu Thr Ile Cys Ala Ala Gly Ile Gly Gly Thr Phe Gln  
 20 25 30  
 Phe Gly Tyr Asn Leu Ser Ile Ile Asn Ala Pro Thr Leu His Ile  
 35 40 45  
 Gln Glu Phe Thr Asn Glu Thr Trp Gln Ala Arg Thr Gly Glu Pro  
 50 55 60  
 Leu Pro Asp His Leu Val Leu Leu Met Trp Ser Leu Ile Val Ser  
 65 70 75  
 Leu Tyr Pro Leu Gly Gly Leu Phe Gly Ala Leu Leu Ala Gly Pro  
 80 85 90  
 Leu Ala Ile Thr Leu Gly Arg Lys Lys Ser Leu Leu Val Asn Asn  
 95 100 105  
 Ile Phe Val Val Ser Ala Ala Ile Leu Phe Gly Phe Ser Arg Lys

Ala Gly Ser Phe	110	Glu Met Ile Met Leu	115	Gly Arg Leu Leu Val	120
Val Asn Ala Gly	125	Val Ser Met Asn Ile	130	Gln Pro Met Tyr Leu	135
Glu Ser Ala Pro	140	Lys Glu Leu Arg Gly	145	Ala Val Ala Met Ser	150
Ala Ile Phe Thr	155	Ala Leu Gly Ile Val	160	Met Gly Gln Val Val	165
Leu Arg Glu Leu	170	Leu Gly Gly Pro Gln	175	Ala Trp Pro Leu Leu	180
Ala Ser Cys Leu	185	Val Pro Gly Ala Leu	190	Gln Leu Ala Ser Leu	195
Leu Leu Pro Glu	200	Ser Pro Arg Tyr Leu	205	Leu Ile Asp Cys Gly	210
Thr Glu Ala Cys	215	Leu Ala Glu Thr Gly	220	Ser Arg Leu Ser Arg	225
Glu Cys Cys Gly	230	Cys Ser	235		240
	245				

&lt;210&gt; 14

&lt;211&gt; 436

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5674114CD1

&lt;400&gt; 14

Met Gly Leu Ala Arg	1	Ala Leu Arg Arg Leu	5	Ser Gly Ala Leu Asp	15
Ser Gly Asp Ser Arg	20	Ala Gly Asp Glu	25	Glu Ala Gly Pro Gly	30
Leu Cys Arg Asn Gly	35	Trp Ala Pro Ala	40	Pro Val Gln Ser Pro Val	45
Gly Arg Arg Arg Gly	50	Arg Phe Val Lys	55	Lys Asp Gly His Cys Asn	60
Val Arg Phe Val Asn	65	Leu Gly Gly Gln	70	Gly Ala Arg Tyr Leu Ser	75
Asp Leu Phe Thr Thr	80	Cys Val Asp Val	85	Arg Trp Arg Trp Met Cys	90
Leu Leu Phe Ser Cys	95	Ser Phe Leu Ala	100	Ser Trp Leu Leu Phe Gly	105
Leu Ala Phe Trp Leu	110	Ile Ala Ser Leu	115	His Gly Asp Leu Ala Ala	120
Pro Pro Pro Pro Ala	125	Pro Cys Phe Ser	130	His Val Ala Ser Phe Leu	135
Ala Ala Phe Leu Phe	140	Ala Leu Glu Thr	145	Gln Thr Ser Ile Gly Tyr	150
Gly Val Arg Ser Val	155	Thr Glu Glu Cys	160	Pro Ala Ala Val Ala Ala	165
Val Val Leu Gln Cys	170	Ile Ala Gly Cys	175	Val Leu Asp Ala Phe Val	180
Val Gly Ala Val Met	185	Ala Lys Met Ala	190	Lys Pro Lys Lys Arg Asn	195
Glu Thr Leu Val Phe	200	Ser Glu Asn Ala	205	Val Val Ala Leu Arg Asp	210
His Arg Leu Cys Leu	215	Met Trp Arg Val	220	Gly Asn Leu Arg Arg Ser	225
His Leu Val Glu Ala	230	His Val Arg Ala	235	Gln Leu Leu Gln Pro Arg	240
Val Thr Pro Glu Gly	245	Glu Tyr Ile Pro	250	Leu Asp His Gln Asp Val	255
Asp Val Gly Phe Asp	260	Gly Gly Thr Asp	265	Arg Ile Phe Leu Val Ser	270
Pro Ile Thr Ile Val		His Glu Ile Asp		Ser Ala Ser Pro Leu Tyr	

	275		280		285
Glu Leu Gly Arg	Ala Glu Leu Ala Arg	Ala Asp Phe Glu Leu	Val		
	290		295		300
Val Ile Leu Glu	Gly Met Val Glu Ala Thr	Ala Met Thr Thr	Gln		
	305		310		315
Cys Arg Ser Ser	Tyr Leu Pro Gly Glu	Leu Leu Trp Gly His	Arg		
	320		325		330
Phe Glu Pro Val	Leu Phe Gln Arg Gly	Ser Gln Tyr Glu Val	Asp		
	335		340		345
Tyr Arg His Phe	His Arg Thr Tyr Glu	Val Pro Gly Thr Pro	Val		
	350		355		360
Cys Ser Ala Lys	Glu Leu Asp Glu Arg	Ala Glu Gln Ala Ser	His		
	365		370		375
Ser Leu Lys Ser	Ser Phe Pro Gly Ser	Leu Thr Ala Phe Cys	Tyr		
	380		385		390
Glu Asn Glu Leu	Ala Leu Ser Cys Cys	Gln Glu Glu Asp Glu	Asp		
	395		400		405
Asp Glu Thr Glu	Gly Glu Asn Gly Val	Gln Thr Glu Asp Gly	Ala		
	410		415		420
Ala Ser Pro Arg	Val Leu Thr Pro Thr	Leu Ala Leu Thr Leu	Pro		
	425		430		435
Pro					

<210> 15  
 <211> 453  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1254635CD1

<400> 15

Met Leu Lys Met Val	Leu Thr Glu Asn Pro	Asn Gln Glu Ile Ala	
1	5	10	15
Thr Ser Leu Glu Phe	Leu Leu Leu Gln Asn	Ser Pro Gly Ser Leu	
	20	25	30
Arg Ala Gln Gln Arg	Met Ser Tyr Tyr Gly	Ser Ser Tyr His Ile	
	35	40	45
Ile Asn Ala Asp Ala	Lys Tyr Pro Gly Tyr	Pro Pro Glu His Ile	
	50	55	60
Ile Ala Glu Lys Arg	Arg Ala Arg Arg Arg	Leu Leu His Lys Asp	
	65	70	75
Gly Ser Cys Asn Val	Tyr Phe Lys His Ile	Phe Gly Glu Trp Gly	
	80	85	90
Ser Tyr Val Val Asp	Ile Phe Thr Thr Leu	Val Asp Thr Lys Trp	
	95	100	105
Arg His Met Phe Val	Ile Phe Ser Leu Ser	Tyr Ile Leu Ser Trp	
	110	115	120
Leu Ile Phe Gly Ser	Val Phe Trp Leu Ile	Ala Phe His His Gly	
	125	130	135
Asp Leu Leu Asn Asp	Pro Asp Ile Thr Pro	Cys Val Asp Asn Val	
	140	145	150
His Ser Phe Thr Gly	Ala Phe Leu Phe Ser	Leu Glu Thr Gln Thr	
	155	160	165
Thr Ile Gly Tyr Gly	Tyr Arg Cys Val Thr	Glu Glu Cys Ser Val	
	170	175	180
Ala Val Leu Met Val	Ile Leu Gln Ser Ile	Leu Ser Cys Ile Ile	
	185	190	195
Asn Thr Phe Ile Ile	Gly Ala Ala Leu Ala	Lys Met Ala Thr Ala	
	200	205	210
Arg Lys Arg Ala Gln	Thr Ile Arg Phe Ser	Tyr Phe Ala Leu Ile	
	215	220	225
Gly Met Arg Asp Gly	Lys Leu Cys Leu Met	Trp Arg Ile Gly Asp	
	230	235	240
Phe Arg Pro Asn His	Val Val Glu Gly Thr	Val Arg Ala Gln Leu	

Leu	Arg	Tyr	Thr	245	Glu	Asp	Ser	Glu	Gly	250	Arg	Met	Thr	Met	Ala	255	Phe
Lys	Asp	Leu	Lys	260	Leu	Val	Asn	Asp	Gln	265	Ile	Ile	Leu	Val	Thr	270	Pro
Val	Thr	Ile	Val	275	His	Glu	Ile	Asp	His	280	Glu	Ser	Pro	Leu	Tyr	285	Ala
Leu	Asp	Arg	Lys	290	Ala	Val	Ala	Lys	Asp	295	Asn	Phe	Glu	Ile	Leu	300	Val
Thr	Phe	Ile	Tyr	305	Thr	Gly	Asp	Ser	Thr	310	Gly	Thr	Ser	His	Gln	315	Ser
Arg	Ser	Ser	Tyr	320	Val	Pro	Arg	Glu	Ile	325	Leu	Trp	Gly	His	Arg	330	Phe
Asn	Asp	Val	Leu	335	Glu	Val	Lys	Arg	Lys	340	Tyr	Tyr	Lys	Val	Asn	345	Cys
Leu	Gln	Phe	Glu	350	Gly	Ser	Val	Glu	Val	355	Tyr	Ala	Pro	Phe	Cys	360	Ser
Ala	Lys	Gln	Leu	365	Asp	Trp	Lys	Asp	Gln	370	Gln	Leu	His	Ile	Glu	375	Lys
Ala	Pro	Pro	Val	380	Arg	Glu	Ser	Cys	Thr	385	Ser	Asp	Thr	Lys	Ala	390	Arg
Arg	Arg	Ser	Phe	395	Ser	Ala	Val	Ala	Ile	400	Val	Ser	Ser	Cys	Glu	405	Asn
Pro	Glu	Glu	Thr	410	Thr	Thr	Ser	Ala	Thr	415	His	Glu	Tyr	Arg	Glu	420	Thr
Pro	Tyr	Gln	Lys	425	Ala	Leu	Leu	Thr	Leu	430	Asn	Arg	Ile	Ser	Val	435	Glu
Ser	Gln	Met		440						445						450	

&lt;210&gt; 16

&lt;211&gt; 299

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1670595CD1

&lt;400&gt; 16

Met	Ala	Ser	Glu	Ser	Ser	Pro	Leu	Leu	Ala	Tyr	Arg	Leu	Leu	Gly
1				5					10					15
Glu	Glu	Gly	Val	Ala	Leu	Pro	Ala	Asn	Gly	Ala	Gly	Gly	Pro	Gly
				20					25					30
Gly	Ala	Ser	Ala	Arg	Lys	Leu	Ser	Thr	Phe	Leu	Gly	Val	Val	Val
				35					40					45
Pro	Thr	Val	Leu	Ser	Met	Phe	Ser	Ile	Val	Val	Phe	Leu	Arg	Ile
				50					55					60
Gly	Phe	Val	Val	Gly	His	Ala	Gly	Leu	Leu	Gln	Ala	Leu	Ala	Met
				65					70					75
Leu	Leu	Val	Ala	Tyr	Phe	Ile	Leu	Ala	Leu	Thr	Val	Leu	Ser	Val
				80					85					90
Cys	Ala	Ile	Ala	Thr	Asn	Gly	Ala	Val	Gln	Gly	Gly	Gly	Ala	Tyr
				95					100					105
Cys	Ile	Leu	Gln	His	Arg	Trp	Thr	Gly	Met	Pro	Gln	Gly	Pro	Val
				110					115					120
Gly	Ser	Gly	Ser	Cys	Pro	Arg	Ala	Thr	Ala	Trp	Asn	Leu	Leu	Tyr
				125					130					135
Gly	Ser	Leu	Leu	Leu	Gly	Leu	Val	Gly	Gly	Val	Cys	Thr	Leu	Gly
				140					145					150
Ala	Gly	Leu	Tyr	Ala	Arg	Ala	Ser	Phe	Leu	Thr	Phe	Leu	Leu	Val
				155					160					165
Ser	Gly	Ser	Leu	Ala	Ser	Val	Leu	Ile	Ser	Phe	Val	Ala	Val	Gly
				170					175					180
Pro	Arg	Asp	Ile	Arg	Leu	Thr	Pro	Arg	Pro	Gly	Pro	Asn	Gly	Ser
				185					190					195
Ser	Leu	Pro	Pro	Arg	Phe	Gly	His	Phe	Thr	Gly	Phe	Asn	Ser	Ser

Thr	Leu	Lys	Asp	Asn	Leu	Gly	Ala	Gly	Tyr	Ala	Glu	Asp	Tyr	Thr
				200					205					210
Thr	Gly	Ala	Val	Met	Asn	Phe	Ala	Ser	Val	Phe	Ala	Val	Leu	Phe
				215					220					225
Asn	Gly	Arg	His	His	Gly	Trp	Gly	Gln	His	Val	Arg	Gly	Ala	Glu
				230					235					240
Gly	Pro	Gln	Pro	Gly	Asp	Pro	Ser	Gly	His	Asp	Arg	Arg	Arg	Arg
				245					250					255
Leu	His	Leu	Leu	Arg	Leu	Cys	Pro	Ala	Phe	Leu	Ser	Leu	Gln	Pro
				260					265					270
Pro	Phe	Thr	Gly	Ala	Leu	Met	Leu	Gly	Ala	Arg	Pro	Pro	Leu	
				275					280					285
				290					295					

<210> 17  
 <211> 606  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1859560CD1

<400> 17

Met	Pro	Ser	Ser	Val	Thr	Ala	Leu	Gly	Gln	Ala	Arg	Ser	Ser	Gly
1				5					10					15
Pro	Gly	Met	Ala	Pro	Ser	Ala	Cys	Cys	Cys	Ser	Pro	Ala	Ala	Leu
				20					25					30
Gln	Arg	Arg	Leu	Pro	Ile	Leu	Ala	Trp	Leu	Pro	Ser	Tyr	Ser	Leu
				35					40					45
Gln	Trp	Leu	Lys	Met	Asp	Phe	Val	Ala	Gly	Leu	Ser	Val	Gly	Leu
				50					55					60
Thr	Ala	Ile	Pro	Gln	Ala	Leu	Ala	Tyr	Ala	Glu	Val	Ala	Gly	Leu
				65					70					75
Pro	Pro	Gln	Tyr	Gly	Leu	Tyr	Ser	Ala	Phe	Met	Gly	Cys	Phe	Val
				80					85					90
Tyr	Phe	Phe	Leu	Gly	Thr	Ser	Arg	Asp	Val	Thr	Leu	Gly	Pro	Thr
				95					100					105
Ala	Ile	Met	Ser	Leu	Leu	Val	Ser	Phe	Tyr	Thr	Phe	His	Glu	Pro
				110					115					120
Ala	Tyr	Ala	Val	Leu	Leu	Ala	Phe	Leu	Ser	Gly	Cys	Ile	Gln	Leu
				125					130					135
Ala	Met	Gly	Val	Leu	Arg	Leu	Gly	Phe	Leu	Leu	Asp	Phe	Ile	Ser
				140					145					150
Tyr	Pro	Val	Ile	Lys	Gly	Phe	Thr	Ser	Ala	Ala	Ala	Val	Thr	Ile
				155					160					165
Gly	Phe	Gly	Gln	Ile	Lys	Asn	Leu	Leu	Gly	Leu	Gln	Asn	Ile	Pro
				170					175					180
Arg	Pro	Phe	Phe	Leu	Gln	Val	Tyr	His	Thr	Phe	Leu	Arg	Ile	Ala
				185					190					195
Glu	Thr	Arg	Val	Gly	Asp	Ala	Val	Leu	Gly	Leu	Val	Cys	Met	Leu
				200					205					210
Leu	Leu	Leu	Val	Leu	Lys	Leu	Met	Arg	Asp	His	Val	Pro	Pro	Val
				215					220					225
His	Pro	Glu	Met	Pro	Pro	Gly	Val	Arg	Leu	Ser	Arg	Gly	Leu	Val
				230					235					240
Trp	Ala	Ala	Thr	Thr	Ala	Arg	Asn	Ala	Leu	Val	Val	Ser	Phe	Ala
				245					250					255
Ala	Leu	Val	Ala	Tyr	Ser	Phe	Glu	Val	Thr	Gly	Tyr	Gln	Pro	Phe
				260					265					270
Ile	Leu	Thr	Gly	Glu	Thr	Ala	Glu	Gly	Leu	Pro	Pro	Val	Arg	Ile
				275					280					285
Pro	Pro	Phe	Ser	Val	Thr	Thr	Ala	Asn	Gly	Thr	Ile	Ser	Phe	Thr
				290					295					300
Glu	Met	Val	Gln	Asp	Met	Gly	Ala	Gly	Leu	Ala	Val	Val	Pro	Leu
				305					310					315
Met	Gly	Leu	Leu	Glu	Ser	Ile	Ala	Val	Ala	Lys	Ala	Phe	Ala	Ser

	320		325		330
Gln Asn Asn Tyr Arg	Ile Asp Ala Asn	Gln Glu Leu Leu Ala	Ile		
	335		340		345
Gly Leu Thr Asn Met	Leu Gly Ser Leu	Val Ser Ser Tyr Pro	Val		
	350		355		360
Thr Gly Ser Phe Gly	Arg Thr Ala Val	Asn Ala Gln Ser Gly	Val		
	365		370		375
Cys Thr Pro Ala Gly	Gly Leu Val Thr	Gly Val Leu Val Leu	Leu		
	380		385		390
Ser Leu Asp Tyr Leu	Thr Ser Leu Phe	Tyr Tyr Ile Pro Lys	Ser		
	395		400		405
Ala Leu Ala Ala Val	Ile Ile Met Ala	Val Ala Pro Leu Phe	Asp		
	410		415		420
Thr Lys Ile Phe Arg	Thr Leu Trp Arg	Val Lys Arg Leu Asp	Leu		
	425		430		435
Leu Pro Leu Cys Val	Thr Phe Leu Leu	Cys Phe Trp Glu Val	Gln		
	440		445		450
Tyr Gly Ile Leu Ala	Gly Ala Leu Val	Ser Leu Leu Met Leu	Leu		
	455		460		465
His Ser Ala Ala Arg	Pro Glu Thr Lys	Val Ser Glu Gly Pro	Val		
	470		475		480
Leu Val Leu Gln Pro	Ala Ser Gly Leu	Ser Phe Pro Ala Met	Glu		
	485		490		495
Ala Leu Arg Glu Glu	Ile Leu Ser Arg	Ala Leu Glu Val Ser	Pro		
	500		505		510
Pro Arg Cys Leu Val	Leu Glu Cys Thr	His Val Cys Ser Ile	Asp		
	515		520		525
Tyr Thr Val Val Leu	Gly Leu Gly Glu	Leu Leu Gln Asp Phe	Gln		
	530		535		540
Lys Gln Gly Val Ala	Leu Ala Phe Val	Gly Leu Gln Val Pro	Val		
	545		550		555
Leu Arg Val Leu Leu	Ser Ala Asp Leu	Lys Gly Phe Gln Tyr	Phe		
	560		565		570
Ser Thr Leu Glu Glu	Ala Glu Lys His	Leu Arg Gln Glu Pro	Gly		
	575		580		585
Thr Gln Pro Tyr Asn	Ile Arg Glu Asp	Ser Ile Leu Asp Gln	Lys		
	590		595		600
Val Ala Leu Leu Lys	Ala				
	605				

&lt;210&gt; 18

&lt;211&gt; 324

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5530164CD1

&lt;400&gt; 18

Met Ser Val Glu Asp	Gly Gly Met Pro	Gly Leu Gly Arg Pro	Arg
1	5	10	15
Gln Ala Arg Trp Thr	Leu Met Leu Leu	Leu Ser Thr Ala Met	Tyr
	20	25	30
Gly Ala His Ala Pro	Leu Leu Ala Leu	Cys His Val Asp Gly	Arg
	35	40	45
Val Pro Phe Arg Pro	Ser Ser Ala Val	Leu Leu Thr Glu Leu	Thr
	50	55	60
Lys Leu Leu Leu Cys	Ala Phe Ser Leu	Leu Val Gly Trp Gln	Ala
	65	70	75
Trp Pro Gln Gly Pro	Pro Pro Trp Arg	Gln Ala Ala Pro Phe	Ala
	80	85	90
Leu Ser Ala Leu Leu	Tyr Gly Ala Asn	Asn Asn Leu Val Ile	Tyr
	95	100	105
Leu Gln Arg Tyr Met	Asp Pro Ser Thr	Tyr Gln Val Leu Ser	Asn
	110	115	120
Leu Lys Ile Gly Ser	Thr Ala Val Leu	Tyr Cys Leu Cys Leu	Arg

His Arg Leu Ser	125	Val Arg Gln Gly Leu	130	Ala Leu Leu Leu Leu	135
	140		145		150
Ala Ala Gly Ala	Cys Tyr Ala Ala Gly	Gly Leu Gln Val Pro	Gly		155
	155		160		165
Asn Thr Leu Pro	Ser Pro Pro Pro Ala	Ala Ala Ala Ser Pro	Met		170
	170		175		180
Pro Leu His Ile	Thr Pro Leu Gly Leu	Leu Leu Leu Ile Leu	Tyr		185
	185		190		195
Cys Leu Ile Ser	Gly Leu Ser Ser Val	Tyr Thr Glu Leu Leu	Met		200
	200		205		210
Lys Arg Gln Arg	Leu Pro Leu Ala Leu	Gln Asn Leu Phe Leu	Tyr		215
	215		220		225
Thr Phe Gly Val	Leu Leu Asn Leu Gly	Leu His Ala Gly Gly	Gly		230
	230		235		240
Ser Gly Pro Gly	Leu Leu Glu Gly Phe	Ser Gly Trp Ala Ala	Leu		245
	245		250		255
Val Val Leu Ser	Gln Ala Leu Asn Gly	Leu Leu Met Ser Ala	Val		260
	260		265		270
Met Lys His Gly	Ser Ser Ile Thr Arg	Leu Phe Val Val Ser	Cys		275
	275		280		285
Ser Leu Val Val	Asn Ala Val Leu Ser	Ala Val Leu Leu Arg	Leu		290
	290		295		300
Gln Leu Thr Ala	Ala Phe Phe Leu Ala	Thr Leu Leu Ile Gly	Leu		305
	305		310		315
Ala Met Arg Leu	Tyr Tyr Gly Ser Arg				320

&lt;210&gt; 19

&lt;211&gt; 445

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 139115CD1

&lt;400&gt; 19

Met Thr Leu Thr Gly	Pro Leu Thr Thr	Gln Tyr Val Tyr Arg	Arg
1	5	10	15
Ile Trp Glu Glu Thr	Gly Asn Tyr Thr	Phe Ser Ser Asp Ser	Asn
	20	25	30
Ile Ser Glu Cys Glu	Lys Asn Lys Ser	Ser Pro Ile Phe Ala	Phe
	35	40	45
Gln Glu Glu Val Gln	Lys Lys Val Ser	Arg Phe Asn Leu Gln	Met
	50	55	60
Asp Ile Ser Gly Leu	Ile Pro Gly Leu	Val Ser Thr Phe Ile	Leu
	65	70	75
Leu Ser Ile Ser Asp	His Tyr Gly Arg	Lys Phe Pro Met Ile	Leu
	80	85	90
Ser Ser Val Gly Ala	Leu Ala Thr Ser	Val Trp Leu Cys Leu	Leu
	95	100	105
Cys Tyr Phe Ala Phe	Pro Phe Gln Leu	Leu Ile Ala Ser Thr	Phe
	110	115	120
Ile Gly Ala Phe Cys	Gly Asn Tyr Thr	Thr Phe Trp Gly Ala	Cys
	125	130	135
Phe Ala Tyr Ile Val	Asp Gln Cys Lys	Glu His Lys Gln Lys	Thr
	140	145	150
Ile Arg Ile Ala Ile	Ile Asp Phe Leu	Leu Gly Leu Val Thr	Gly
	155	160	165
Leu Thr Gly Leu Ser	Ser Gly Tyr Phe	Ile Arg Glu Leu Gly	Phe
	170	175	180
Glu Trp Ser Phe Leu	Ile Ile Ala Val	Ser Leu Ala Val Asn	Leu
	185	190	195
Ile Tyr Ile Leu Phe	Phe Leu Gly Asp	Pro Val Lys Glu Cys	Ser
	200	205	210
Ser Gln Asn Val Thr	Met Ser Cys Ser	Glu Gly Phe Lys Asn	Leu

Phe	Tyr	Arg	Thr	215	Tyr	Met	Leu	Phe	Lys	220	Asn	Ala	Ser	Gly	Lys	225	Arg
				230						235							240
Arg	Phe	Leu	Leu	245	Cys	Leu	Leu	Leu	Phe	250	Thr	Val	Ile	Thr	Tyr	Phe	
Phe	Val	Val	Ile	260	Gly	Ile	Ala	Pro	Ile	265	Phe	Ile	Leu	Tyr	Glu	Leu	
Asp	Ser	Pro	Leu	275	Cys	Trp	Asn	Glu	Val	280	Phe	Ile	Gly	Tyr	Gly	Ser	
Ala	Leu	Gly	Ser	290	Ala	Ser	Phe	Leu	Thr	295	Ser	Phe	Leu	Gly	Ile	Trp	
Leu	Phe	Ser	Tyr	305	Cys	Met	Glu	Asp	Ile	310	His	Met	Ala	Phe	Ile	Gly	
Ile	Phe	Thr	Thr	320	Met	Thr	Gly	Met	Ala	325	Met	Thr	Ala	Phe	Ala	Ser	
Thr	Thr	Leu	Met	335	Met	Phe	Leu	Ala	Arg	340	Val	Pro	Phe	Leu	Phe	Thr	
Ile	Val	Pro	Phe	350	Ser	Val	Leu	Arg	Ser	355	Met	Leu	Ser	Lys	Val	Val	
Arg	Ser	Thr	Glu	365	Gln	Gly	Thr	Leu	Phe	370	Ala	Cys	Ile	Ala	Phe	Leu	
Glu	Thr	Leu	Gly	380	Gly	Val	Thr	Ala	Val	385	Ser	Thr	Phe	Asn	Gly	Ile	
Tyr	Ser	Ala	Thr	395	Val	Ala	Trp	Tyr	Pro	400	Gly	Phe	Thr	Phe	Leu	Leu	
Ser	Ala	Gly	Leu	410	Leu	Leu	Leu	Pro	Ala	415	Ile	Ser	Leu	Cys	Val	Val	
Lys	Cys	Thr	Ser	425	Trp	Asn	Glu	Gly	Ser	430	Tyr	Glu	Leu	Leu	Ile	Gln	
Glu	Glu	Ser	Ser	440	Glu	Asp	Ala	Ser	Asp	445	Arg						

&lt;210&gt; 20

&lt;211&gt; 337

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1702940CD1

&lt;400&gt; 20

Met	Asn	Pro	Glu	Ser	Ser	Ile	Phe	Ile	Glu	Asp	Tyr	Leu	Lys	Tyr			
1				5					10					15			
Phe	Gln	Asp	Gln	Val	Ser	Arg	Glu	Asn	Leu	Leu	Gln	Leu	Leu	Thr			
				20					25					30			
Asp	Asp	Glu	Ala	Trp	Asn	Gly	Phe	Val	Ala	Ala	Ala	Glu	Leu	Pro			
				35					40					45			
Arg	Asp	Glu	Ala	Asp	Glu	Leu	Arg	Lys	Ala	Leu	Asn	Lys	Leu	Ala			
				50					55					60			
Ser	His	Met	Val	Met	Lys	Asp	Lys	Asn	Arg	His	Asp	Lys	Asp	Gln			
				65					70					75			
Gln	His	Arg	Gln	Trp	Phe	Leu	Lys	Glu	Phe	Pro	Arg	Leu	Lys	Arg			
				80					85					90			
Glu	Leu	Glu	Asp	His	Ile	Arg	Lys	Leu	Arg	Ala	Leu	Ala	Glu	Glu			
				95					100					105			
Val	Glu	Gln	Val	His	Arg	Gly	Thr	Thr	Ile	Ala	Asn	Val	Val	Ser			
				110					115					120			
Asn	Ser	Val	Gly	Thr	Thr	Ser	Gly	Ile	Leu	Thr	Leu	Leu	Gly	Leu			
				125					130					135			
Gly	Leu	Ala	Pro	Phe	Thr	Glu	Gly	Ile	Ser	Phe	Val	Leu	Leu	Asp			
				140					145					150			
Thr	Gly	Met	Gly	Leu	Gly	Ala	Ala	Ala	Ala	Val	Ala	Gly	Ile	Thr			
				155					160					165			
Cys	Ser	Val	Val	Glu	Leu	Val	Asn	Lys	Leu	Arg	Ala	Arg	Ala	Gln			
				170					175					180			
Ala	Arg	Asn	Leu	Asp	Gln	Ser	Gly	Thr	Asn	Val	Ala	Lys	Val	Met			



Lys	Glu	Phe	Val	185	Gly	Gly	Asn	Thr	Pro	190	Asn	Val	Leu	Thr	Leu	195	Val
Asp	Asn	Trp	Tyr	200	Gln	Val	Thr	Gln	Gly	205	Ile	Gly	Arg	Asn	Ile	210	Arg
Ala	Ile	Arg	Arg	215	Ala	Arg	Ala	Asn	Pro	220	Gln	Leu	Gly	Ala	Tyr	225	Ala
Pro	Pro	Pro	His	230	Val	Ile	Gly	Arg	Ile	235	Ser	Ala	Glu	Gly	Gly	240	Glu
Gln	Val	Glu	Arg	245	Val	Val	Glu	Gly	Pro	250	Ala	Gln	Ala	Met	Ser	255	Arg
Gly	Thr	Met	Ile	260	Val	Gly	Ala	Ala	Thr	265	Gly	Gly	Ile	Leu	Leu	270	Leu
Leu	Asp	Val	Val	275	Ser	Leu	Ala	Tyr	Glu	280	Ser	Lys	His	Leu	Leu	285	Glu
Gly	Ala	Lys	Ser	290	Glu	Ser	Ala	Glu	Glu	295	Leu	Lys	Lys	Arg	Ala	300	Gln
Glu	Leu	Glu	Gly	305	Lys	Leu	Asn	Phe	Leu	310	Lys	Ile	His	Glu	Met	315	Met
Leu	Gln	Pro	Gly	320	Gln	Asp	Gln			325						330	
				335													

<210> 21  
 <211> 273  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1703342CD1

<400> 21

Met	Ala	Thr	Trp	Asp	Glu	Lys	Ala	Val	Thr	Arg	Arg	Ala	Lys	Val
1				5					10					15
Ala	Pro	Ala	Glu	Arg	Met	Ser	Lys	Phe	Leu	Arg	His	Phe	Thr	Val
				20					25					30
Val	Gly	Asp	Asp	Tyr	His	Ala	Trp	Asn	Ile	Asn	Tyr	Lys	Lys	Trp
				35					40					45
Glu	Asn	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Gln	Pro	Pro	Pro	Thr
				50					55					60
Pro	Val	Ser	Gly	Glu	Glu	Gly	Arg	Ala	Ala	Ala	Pro	Asp	Val	Ala
				65					70					75
Pro	Ala	Pro	Gly	Pro	Ala	Pro	Arg	Ala	Pro	Leu	Asp	Phe	Arg	Gly
				80					85					90
Met	Leu	Arg	Lys	Leu	Phe	Ser	Ser	His	Arg	Phe	Gln	Val	Ile	Ile
				95					100					105
Ile	Cys	Leu	Val	Val	Leu	Asp	Ala	Leu	Leu	Val	Leu	Ala	Glu	Leu
				110					115					120
Ile	Leu	Asp	Leu	Lys	Ile	Ile	Gln	Pro	Asp	Lys	Asn	Asn	Tyr	Ala
				125					130					135
Ala	Met	Val	Phe	His	Tyr	Met	Ser	Ile	Thr	Ile	Leu	Val	Phe	Phe
				140					145					150
Met	Met	Glu	Ile	Ile	Phe	Lys	Leu	Phe	Val	Phe	Arg	Leu	Glu	Phe
				155					160					165
Phe	His	His	Lys	Phe	Glu	Ile	Leu	Asp	Ala	Val	Val	Val	Val	Val
				170					175					180
Ser	Phe	Ile	Leu	Asp	Ile	Val	Leu	Leu	Phe	Gln	Glu	His	Gln	Phe
				185					190					195
Glu	Ala	Leu	Gly	Leu	Leu	Ile	Leu	Leu	Arg	Leu	Trp	Arg	Val	Ala
				200					205					210
Arg	Ile	Ile	Asn	Gly	Ile	Ile	Ile	Ser	Val	Lys	Thr	Arg	Ser	Glu
				215					220					225
Arg	Gln	Leu	Leu	Arg	Leu	Lys	Gln	Met	Asn	Val	Gln	Leu	Ala	Ala
				230					235					240
Lys	Ile	Gln	His	Leu	Glu	Phe	Ser	Cys	Ser	Glu	Lys	Glu	Gln	Glu
				245					250					255
Ile	Glu	Arg	Leu	Asn	Lys	Leu	Leu	Arg	Gln	His	Gly	Leu	Leu	Gly

Glu Val Asn 260 265 270  
 <210> 22  
 <211> 710  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1727529CD1  
 <400> 22  
 Met Gly Gly Lys Gln Arg Asp Glu Asp Asp Glu Ala Tyr Gly Lys  
 1 5 10 15  
 Pro Val Lys Tyr Asp Pro Ser Phe Arg Gly Pro Ile Lys Asn Arg  
 20 25 30  
 Ser Cys Thr Asp Val Ile Cys Cys Val Leu Phe Leu Leu Phe Ile  
 35 40 45  
 Leu Gly Tyr Ile Val Val Gly Ile Val Ala Trp Leu Tyr Gly Asp  
 50 55 60  
 Pro Arg Gln Val Leu Tyr Pro Arg Asn Ser Thr Gly Ala Tyr Cys  
 65 70 75  
 Gly Met Gly Glu Asn Lys Asp Lys Pro Tyr Leu Leu Tyr Phe Asn  
 80 85 90  
 Ile Phe Ser Cys Ile Leu Ser Ser Asn Ile Ile Ser Val Ala Glu  
 95 100 105  
 Asn Gly Leu Gln Cys Pro Thr Pro Gln Val Cys Val Ser Ser Cys  
 110 115 120  
 Pro Glu Asp Pro Trp Thr Val Gly Lys Asn Glu Phe Ser Gln Thr  
 125 130 135  
 Val Gly Glu Val Phe Tyr Thr Lys Asn Arg Asn Phe Cys Leu Pro  
 140 145 150  
 Gly Val Pro Trp Asn Met Thr Val Ile Thr Ser Leu Gln Gln Glu  
 155 160 165  
 Leu Cys Pro Ser Phe Leu Leu Pro Ser Ala Pro Ala Leu Gly Arg  
 170 175 180  
 Cys Phe Pro Trp Thr Asn Ile Thr Pro Pro Ala Leu Pro Gly Ile  
 185 190 195  
 Thr Asn Asp Thr Thr Ile Gln Gln Gly Ile Ser Gly Leu Ile Asp  
 200 205 210  
 Ser Leu Asn Ala Arg Asp Ile Ser Val Lys Ile Phe Glu Asp Phe  
 215 220 225  
 Ala Gln Ser Trp Tyr Trp Ile Leu Val Ala Leu Gly Val Ala Leu  
 230 235 240  
 Val Leu Ser Leu Leu Phe Ile Leu Leu Leu Arg Leu Val Ala Gly  
 245 250 255  
 Pro Leu Val Leu Val Leu Ile Leu Gly Val Leu Gly Val Leu Ala  
 260 265 270  
 Tyr Gly Ile Tyr Tyr Cys Trp Glu Glu Tyr Arg Val Leu Arg Asp  
 275 280 285  
 Lys Gly Ala Ser Ile Ser Gln Leu Gly Phe Thr Thr Asn Leu Ser  
 290 295 300  
 Ala Tyr Gln Ser Val Gln Glu Thr Trp Leu Ala Ala Leu Ile Val  
 305 310 315  
 Leu Ala Val Leu Glu Ala Ile Leu Leu Leu Val Leu Ile Phe Leu  
 320 325 330  
 Arg Gln Arg Ile Arg Ile Ala Ile Ala Leu Leu Lys Glu Ala Ser  
 335 340 345  
 Lys Ala Val Gly Gln Met Met Ser Thr Met Phe Tyr Pro Leu Val  
 350 355 360  
 Thr Phe Val Leu Leu Leu Ile Cys Ile Ala Tyr Trp Ala Met Thr  
 365 370 375  
 Ala Leu Tyr Leu Ala Thr Ser Gly Gln Pro Gln Tyr Val Leu Trp  
 380 385 390  
 Ala Ser Asn Ile Ser Ser Pro Gly Cys Glu Lys Val Pro Ile Asn

Thr	Ser	Cys	Asn	395	Thr	Ala	His	Leu	400	Val	Asn	Ser	Ser	Cys	405
				410					415						420
Gly	Leu	Met	Cys	425	Val	Phe	Gln	Gly	430	Ser	Ser	Lys	Gly	Leu	435
				440					445						450
Gln	Arg	Ser	Val	455	Phe	Asn	Leu	Gln	460	Tyr	Gly	Val	Leu	Gly	465
				470					475						480
Phe	Trp	Thr	Leu	485	Asn	Trp	Val	Leu	490	Leu	Gly	Gln	Cys	Val	495
				500					505						510
Ala	Gly	Ala	Phe	515	Ala	Ser	Phe	Tyr	520	Ala	Phe	His	Lys	Pro	525
				530					535						540
Asp	Ile	Pro	Thr	545	Phe	Pro	Leu	Ile	550	Ala	Phe	Ile	Arg	Thr	555
				560					565						570
Arg	Tyr	His	Thr	575	Gly	Ser	Leu	Ala	580	Gly	Ala	Leu	Ile	Leu	585
				590					595						600
Leu	Val	Gln	Ile	605	Ala	Arg	Val	Ile	610	Glu	Tyr	Ile	Asp	His	615
				620					625						630
Leu	Arg	Gly	Val	635	Gln	Asn	Pro	Val	640	Arg	Cys	Ile	Met	Cys	645
				650					655						660
Phe	Lys	Cys	Cys	665	Leu	Trp	Cys	Leu	670	Lys	Phe	Ile	Lys	Phe	675
				680					685						690
Asn	Arg	Asn	Ala	695	Tyr	Ile	Met	Ile	700	Ile	Tyr	Gly	Lys	Asn	705
				710											
Cys	Val	Ser	Ala		Lys	Asn	Ala	Phe		Leu	Leu	Met	Arg	Asn	
Val	Arg	Val	Val		Val	Leu	Asp	Lys		Thr	Asp	Leu	Leu	Leu	
Phe	Gly	Lys	Leu		Leu	Val	Val	Gly		Val	Gly	Val	Leu	Ser	
Phe	Phe	Phe	Ser		Gly	Arg	Ile	Pro		Leu	Gly	Lys	Asp	Phe	
Ser	Pro	His	Leu		Asn	Tyr	Tyr	Trp		Pro	Ile	Met	Thr	Ser	
Leu	Gly	Ala	Tyr		Val	Ile	Ala	Ser		Phe	Phe	Ser	Val	Phe	
Met	Cys	Val	Asp		Thr	Leu	Phe	Leu		Phe	Leu	Glu	Asp	Leu	
Arg	Asn	Asn	Gly		Ser	Leu	Asp	Arg		Tyr	Tyr	Met	Ser	Lys	
Leu	Leu	Lys	Ile		Leu	Gly	Lys	Lys		Glu	Ala	Pro	Pro	Asp	
Lys	Lys	Arg	Lys		Lys										

<210> 23  
 <211> 476  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2289333CD1

<400> 23  
 Glu Gln Asn Phe Asp Gly Thr Ser Asp Glu Glu His Glu Gln Glu  
 1 5 10 15  
 Leu Leu Pro Val Gln Lys His Tyr Gln Leu Asp Asp Gln Glu Gly  
 20 25 30  
 Ile Ser Phe Val Gln Thr Leu Met His Leu Leu Lys Gly Asn Ile  
 35 40 45  
 Gly Thr Gly Leu Leu Gly Leu Pro Leu Ala Ile Lys Asn Ala Gly  
 50 55 60  
 Ile Val Leu Gly Pro Ile Ser Leu Val Phe Ile Gly Ile Ile Ser  
 65 70 75  
 Val His Cys Met His Ile Leu Val Arg Cys Ser His Phe Leu Cys  
 80 85 90  
 Leu Arg Phe Lys Lys Ser Thr Leu Gly Tyr Ser Asp Thr Val Ser

Phe	Ala	Met	Glu	Val	Ser	Pro	Trp	Ser	Cys	Leu	Gln	Lys	Gln	Ala
Ala	Trp	Gly	Arg	Ser	Val	Val	Asp	Phe	Phe	Leu	Val	Ile	Thr	Gln
Leu	Gly	Phe	Cys	Ser	Val	Tyr	Ile	Val	Phe	Leu	Ala	Glu	Asn	Val
Lys	Gln	Val	His	Glu	Gly	Phe	Leu	Glu	Ser	Lys	Val	Phe	Ile	Ser
Asn	Ser	Thr	Asn	Ser	Ser	Asn	Pro	Cys	Glu	Arg	Arg	Ser	Val	Asp
Leu	Arg	Ile	Tyr	Met	Leu	Cys	Phe	Leu	Pro	Phe	Ile	Ile	Leu	Leu
Val	Phe	Ile	Arg	Glu	Leu	Lys	Asn	Leu	Phe	Val	Leu	Ser	Phe	Leu
Ala	Asn	Val	Ser	Met	Ala	Val	Ser	Leu	Val	Ile	Ile	Tyr	Gln	Tyr
Val	Val	Arg	Asn	Met	Pro	Asp	Pro	His	Asn	Leu	Pro	Ile	Val	Ala
Gly	Trp	Lys	Lys	Tyr	Pro	Leu	Phe	Phe	Gly	Thr	Ala	Val	Phe	Ala
Phe	Glu	Gly	Ile	Gly	Val	Val	Leu	Pro	Leu	Glu	Asn	Gln	Met	Lys
Glu	Ser	Lys	Arg	Phe	Pro	Gln	Ala	Leu	Asn	Ile	Gly	Met	Gly	Ile
Val	Thr	Thr	Leu	Tyr	Val	Thr	Leu	Ala	Thr	Leu	Gly	Tyr	Met	Cys
Phe	His	Asp	Glu	Ile	Lys	Gly	Ser	Ile	Thr	Leu	Asn	Leu	Pro	Gln
Asp	Val	Trp	Leu	Tyr	Gln	Ser	Val	Lys	Ile	Leu	Tyr	Ser	Phe	Gly
Ile	Phe	Val	Thr	Tyr	Ser	Ile	Gln	Phe	Tyr	Val	Pro	Ala	Glu	Ile
Ile	Ile	Pro	Gly	Ile	Thr	Ser	Lys	Phe	His	Thr	Lys	Trp	Lys	Gln
Ile	Cys	Glu	Phe	Gly	Ile	Arg	Ser	Phe	Leu	Val	Ser	Ile	Thr	Cys
Ala	Gly	Ala	Ile	Leu	Ile	Pro	Arg	Leu	Asp	Ile	Val	Ile	Ser	Phe
Val	Gly	Ala	Val	Ser	Ser	Ser	Thr	Leu	Ala	Leu	Ile	Leu	Pro	Pro
Leu	Val	Glu	Ile	Leu	Thr	Phe	Ser	Lys	Glu	His	Tyr	Asn	Ile	Trp
Met	Val	Leu	Lys	Asn	Ile	Ser	Ile	Ala	Phe	Thr	Gly	Val	Val	Gly
Phe	Leu	Leu	Gly	Thr	Tyr	Ile	Thr	Val	Glu	Glu	Ile	Ile	Tyr	Pro
Thr	Pro	Lys	Val	Val	Ala	Gly	Thr	Pro	Gln	Ser	Pro	Phe	Leu	Asn
Leu	Asn	Ser	Thr	Cys	Leu	Thr	Ser	Gly	Leu	Lys				

<210> 24  
 <211> 237  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2720354CD1

<400> 24  
 Met Gly Leu Thr Phe Ile Asn Ala Leu Val Phe Gly Val Gln Gly  
 1 5 10 15  
 Asn Thr Leu Arg Ala Leu Gly His Asp Ser Pro Leu Asn Gln Phe  
 20 25 30  
 Leu Ala Gly Ala Ala Ala Gly Ala Ile Gln Cys Val Ile Cys Cys

Pro Met Glu Leu	35	Ala Lys Thr Arg Leu	40	Gln Leu Gln Asp Ala	45
	50		55		60
Pro Ala Arg Thr	65	Tyr Lys Gly Ser Leu	70	Cys Leu Ala Gln	75
	80		85		90
Tyr Gly His Glu	95	Gly Leu Arg Gly Val	100	Asn Arg Gly Met Val	105
	110		115		120
Thr Leu Leu Arg	125	Glu Thr Pro Ser Phe	130	Gly Val Tyr Phe Leu	135
	140		145		150
Tyr Asp Ala Leu	155	Thr Arg Ala Leu Gly	160	Cys Glu Pro Gly Asp	165
	170		175		180
Leu Leu Val Pro	185	Lys Leu Leu Leu Ala	190	Gly Gly Thr Ser Gly	195
	200		205		210
Val Ser Trp Leu	215	Ser Thr Tyr Pro Val	220	Asp Val Val Lys Ser	225
	230		235		
Leu Gln Ala Asp		Gly Leu Arg Gly Ala		Pro Arg Tyr Arg Gly	
Leu Asp Cys Val		His Gln Ser Tyr Arg		Ala Glu Gly Trp Arg	
Phe Thr Arg Gly		Leu Ala Ser Thr Leu		Leu Arg Ala Phe Pro	
Asn Ala Ala Thr		Phe Ala Thr Val Thr		Val Val Leu Thr Tyr	
Arg Gly Glu Glu		Ala Gly Pro Glu Gly		Glu Ala Val Pro Ala	
Pro Ala Gly Pro		Ala Leu Ala Gln Pro		Ser Ser Leu	

&lt;210&gt; 25

&lt;211&gt; 345

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3038193CD1

&lt;400&gt; 25

Met Arg Leu Leu	5	Glu Arg Met Arg Lys	10	Asp Trp Phe Met Val	15
	20		25		30
Ile Val Leu Ala	35	Ile Ala Gly Ala Lys	40	Leu Glu Pro Ser Ile	45
	50		55		60
Val Asn Gly Gly	65	Pro Leu Lys Pro Glu	70	Ile Thr Val Ser Tyr	75
	80		85		90
Ala Val Ala Thr	95	Ile Phe Phe Asn Ser	100	Gly Leu Ser Leu Lys	105
	110		115		120
Glu Glu Leu Thr	125	Ser Ala Leu Val His	130	Leu Lys Leu His Leu	135
	140		145		150
Ile Gln Ile Phe	155	Thr Leu Ala Phe Phe	160	Pro Ala Thr Ile Trp	165
	170		175		180
Phe Leu Gln Leu	185	Leu Ser Ile Thr Pro	190	Ile Asn Glu Trp Leu	195
	200		205		210
Lys Gly Leu Gln	215	Thr Val Gly Cys Met	220	Pro Pro Pro Val Ser	225
	230		235		
Ala Val Ile Leu		Thr Lys Ala Val Gly		Gly Asn Glu Gly Ile	
Ile Thr Pro Leu		Leu Leu Leu Phe		Leu Gly Ser Ser Ser	
Val Pro Phe Thr		Ser Ile Phe Ser Gln		Leu Phe Met Thr Val	
Val Pro Leu Ile		Ile Gly Gln Ile Val		Arg Arg Tyr Ile Lys	
Trp Leu Glu Arg		Lys Lys Pro Pro Phe		Gly Ala Ile Ser Ser	
Val Leu Leu Met		Ile Ile Tyr Thr Thr		Phe Cys Asp Thr Phe	
Asn Pro Asn Ile		Asp Leu Asp Lys Phe		Ser Leu Val Leu Ile	

				215					220					225
Phe	Ile	Ile	Phe	Ser	Ile	Gln	Leu	Ser	Phe	Met	Leu	Leu	Thr	Phe
				230					235					240
Ile	Phe	Ser	Thr	Arg	Asn	Asn	Ser	Gly	Phe	Thr	Pro	Ala	Asp	Thr
				245					250					255
Val	Ala	Ile	Ile	Phe	Cys	Ser	Thr	His	Lys	Ser	Leu	Thr	Leu	Gly
				260					265					270
Ile	Pro	Met	Leu	Lys	Ile	Val	Phe	Ala	Gly	Tyr	Glu	His	Leu	Ser
				275					280					285
Leu	Ile	Ser	Val	Pro	Leu	Leu	Ile	Tyr	His	Pro	Ala	Gln	Ile	Leu
				290					295					300
Leu	Gly	Ser	Val	Leu	Val	Pro	Thr	Ile	Lys	Ser	Trp	Met	Val	Ser
				305					310					315
Arg	Gln	Lys	Lys	Leu	Leu	Gln	Thr	Arg	Gly	Pro	Leu	Ala	Asn	Leu
				320					325					330
Asn	Asn	Pro	Glu	Gly	Leu	Glu	Tyr	Leu	Ser	Ile	Lys	Phe	Gly	His
				335					340					345

&lt;210&gt; 26

&lt;211&gt; 521

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3460979CD1

&lt;400&gt; 26

Met	Ala	Ala	Leu	Ala	Pro	Val	Gly	Ser	Pro	Ala	Ser	Arg	Gly	Pro
1				5					10					15
Arg	Leu	Ala	Ala	Gly	Leu	Arg	Leu	Leu	Pro	Met	Leu	Gly	Leu	Leu
				20					25					30
Gln	Leu	Leu	Ala	Glu	Pro	Gly	Leu	Gly	Arg	Val	His	His	Leu	Ala
				35					40					45
Leu	Lys	Asp	Asp	Val	Arg	His	Lys	Val	His	Leu	Asn	Thr	Phe	Gly
				50					55					60
Phe	Phe	Lys	Asp	Gly	Tyr	Met	Val	Val	Asn	Val	Ser	Ser	Leu	Ser
				65					70					75
Leu	Asn	Glu	Pro	Glu	Asp	Lys	Asp	Val	Thr	Ile	Gly	Phe	Ser	Leu
				80					85					90
Asp	Arg	Thr	Lys	Asn	Asp	Gly	Phe	Ser	Ser	Tyr	Leu	Asp	Glu	Asp
				95					100					105
Val	Asn	Tyr	Cys	Ile	Leu	Lys	Lys	Gln	Ser	Val	Ser	Val	Thr	Leu
				110					115					120
Leu	Ile	Leu	Asp	Ile	Ser	Arg	Ser	Glu	Val	Arg	Val	Lys	Ser	Pro
				125					130					135
Pro	Glu	Ala	Gly	Thr	Gln	Leu	Pro	Lys	Ile	Ile	Phe	Ser	Arg	Asp
				140					145					150
Glu	Lys	Val	Leu	Gly	Gln	Ser	Gln	Glu	Pro	Asn	Val	Asn	Pro	Ala
				155					160					165
Ser	Ala	Gly	Asn	Gln	Thr	Gln	Lys	Thr	Gln	Asp	Gly	Gly	Lys	Ser
				170					175					180
Lys	Arg	Ser	Thr	Val	Asp	Ser	Lys	Ala	Met	Gly	Glu	Lys	Ser	Phe
				185					190					195
Ser	Val	His	Asn	Asn	Gly	Gly	Ala	Val	Ser	Phe	Gln	Phe	Phe	Phe
				200					205					210
Asn	Ile	Ser	Thr	Asp	Asp	Gln	Glu	Gly	Leu	Tyr	Ser	Leu	Tyr	Phe
				215					220					225
His	Lys	Cys	Leu	Gly	Lys	Glu	Leu	Pro	Ser	Asp	Lys	Phe	Thr	Phe
				230					235					240
Ser	Leu	Asp	Ile	Glu	Ile	Thr	Glu	Lys	Asn	Pro	Asp	Ser	Tyr	Leu
				245					250					255
Ser	Ala	Gly	Glu	Ile	Pro	Leu	Pro	Lys	Leu	Tyr	Ile	Ser	Met	Ala
				260					265					270
Phe	Phe	Phe	Phe	Leu	Ser	Gly	Thr	Ile	Trp	Ile	His	Ile	Leu	Arg
				275					280					285

Lys	Arg	Arg	Asn	Asp	Val	Phe	Lys	Ile	His	Trp	Leu	Met	Ala	Ala	
				290					295					300	
Leu	Pro	Phe	Thr	Lys	Ser	Leu	Ser	Leu	Val	Phe	His	Ala	Ile	Asp	
				305					310					315	
Tyr	His	Tyr	Ile	Ser	Ser	Gln	Gly	Phe	Pro	Ile	Glu	Gly	Trp	Ala	
				320					325					330	
Val	Val	Tyr	Tyr	Ile	Thr	His	Leu	Leu	Lys	Gly	Ala	Leu	Leu	Phe	
				335					340					345	
Ile	Thr	Ile	Ala	Leu	Ile	Gly	Thr	Gly	Trp	Ala	Phe	Ile	Lys	His	
				350					355					360	
Ile	Leu	Ser	Asp	Lys	Asp	Lys	Lys	Ile	Phe	Met	Ile	Val	Ile	Pro	
				365					370					375	
Leu	Gln	Val	Leu	Ala	Asn	Val	Ala	Tyr	Ile	Ile	Ile	Glu	Ser	Thr	
				380					385					390	
Glu	Glu	Gly	Thr	Thr	Glu	Tyr	Gly	Leu	Trp	Lys	Asp	Ser	Leu	Phe	
				395					400					405	
Leu	Val	Asp	Leu	Leu	Cys	Cys	Gly	Ala	Ile	Leu	Phe	Pro	Val	Val	
				410					415					420	
Trp	Ser	Ile	Arg	His	Leu	Gln	Glu	Ala	Ser	Ala	Thr	Asp	Gly	Lys	
				425					430					435	
Ala	Ala	Ile	Asn	Leu	Ala	Lys	Leu	Lys	Leu	Phe	Arg	His	Tyr	Tyr	
				440					445					450	
Val	Leu	Ile	Val	Cys	Tyr	Ile	Tyr	Phe	Thr	Arg	Ile	Ile	Ala	Phe	
				455					460					465	
Leu	Leu	Lys	Leu	Ala	Val	Pro	Phe	Gln	Trp	Lys	Trp	Leu	Tyr	Gln	
				470					475					480	
Leu	Leu	Asp	Glu	Thr	Ala	Thr	Leu	Val	Phe	Phe	Val	Leu	Thr	Gly	
				485					490					495	
Tyr	Lys	Phe	Arg	Pro	Ala	Ser	Asp	Asn	Pro	Tyr	Leu	Gln	Leu	Ser	
				500					505					510	
Gln	Glu	Glu	Glu	Asp	Leu	Glu	Met	Glu	Ser	Val					
				515					520						

&lt;210&gt; 27

&lt;211&gt; 555

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472200CD1

&lt;400&gt; 27

Met	Thr	Leu	Val	Tyr	Phe	Pro	Pro	Ser	Lys	Leu	Gln	Gln	Gln	Gln	
1				5					10					15	
Gln	Pro	Ser	Arg	Ser	Ser	Arg	Leu	Ala	Gln	Gln	Leu	Ala	Gln	Ser	
				20					25					30	
Ser	Trp	Gln	Leu	Ala	Leu	Arg	Phe	Gly	Lys	Arg	Thr	Thr	Ile	His	
				35					40					45	
Gly	Leu	Asp	Arg	Leu	Leu	Ser	Ala	Lys	Ala	Ser	Arg	Trp	Glu	Arg	
				50					55					60	
Phe	Val	Trp	Leu	Cys	Thr	Phe	Val	Ser	Ala	Phe	Leu	Gly	Ala	Val	
				65					70					75	
Tyr	Val	Cys	Leu	Ile	Leu	Ser	Ala	Arg	Tyr	Asn	Ala	Ala	His	Phe	
				80					85					90	
Gln	Thr	Val	Val	Asp	Ser	Thr	Arg	Phe	Pro	Val	Tyr	Arg	Ile	Pro	
				95					100					105	
Phe	Pro	Val	Ile	Thr	Ile	Cys	Asn	Arg	Asn	Arg	Leu	Asn	Trp	Gln	
				110					115					120	
Arg	Leu	Ala	Glu	Ala	Lys	Ser	Arg	Phe	Leu	Ala	Asn	Gly	Ser	Asn	
				125					130					135	
Ser	Ala	Gln	Gln	Glu	Leu	Phe	Glu	Leu	Ile	Val	Gly	Thr	Tyr	Asp	
				140					145					150	
Asp	Ala	Tyr	Phe	Gly	His	Phe	Gln	Ser	Phe	Glu	Arg	Leu	Arg	Asn	
				155					160					165	
Gln	Pro	Thr	Glu	Leu	Leu	Asn	Tyr	Val	Asn	Phe	Ser	Gln	Val	Val	
				170					175					180	

Asp	Phe	Met	Thr	Trp	Arg	Cys	Asn	Glu	Leu	Leu	Ala	Glu	Cys	Leu	
				185					190					195	
Trp	Arg	His	His	Ala	Tyr	Asp	Cys	Cys	Glu	Ile	Arg	Ser	Lys	Arg	
				200					205					210	
Arg	Ser	Lys	Asn	Gly	Leu	Cys	Trp	Ala	Phe	Asn	Ser	Leu	Glu	Thr	
				215					220					225	
Glu	Glu	Gly	Arg	Arg	Met	Gln	Leu	Leu	Asp	Pro	Met	Trp	Pro	Trp	
				230					235					240	
Arg	Thr	Gly	Ser	Ala	Gly	Pro	Met	Ser	Ala	Leu	Ser	Val	Arg	Val	
				245					250					255	
Leu	Ile	Gln	Pro	Ala	Lys	His	Trp	Pro	Gly	His	Arg	Glu	Thr	Asn	
				260					265					270	
Ala	Met	Lys	Gly	Ile	Asp	Val	Met	Val	Thr	Glu	Pro	Phe	Val	Trp	
				275					280					285	
His	Asn	Asn	Pro	Phe	Phe	Val	Ala	Ala	Asn	Thr	Glu	Thr	Thr	Met	
				290					295					300	
Glu	Ile	Glu	Pro	Val	Ile	Tyr	Phe	Tyr	Asp	Asn	Asp	Thr	Arg	Gly	
				305					310					315	
Val	Arg	Ser	Asp	Gln	Arg	Gln	Cys	Val	Phe	Asp	Asp	Glu	His	Asn	
				320					325					330	
Ser	Lys	Asp	Phe	Lys	Ser	Leu	Gln	Gly	Tyr	Val	Tyr	Met	Ile	Glu	
				335					340					345	
Asn	Cys	Gln	Ser	Glu	Cys	His	Gln	Glu	Tyr	Leu	Val	Arg	Tyr	Cys	
				350					355					360	
Asn	Cys	Thr	Met	Asp	Leu	Leu	Phe	Pro	Pro	Asp	Leu	Leu	Ile	Tyr	
				365					370					375	
Ser	His	Asn	Pro	Gly	Glu	Lys	Glu	Phe	Val	Arg	Asn	Gln	Phe	Gln	
				380					385					390	
Gly	Met	Ser	Cys	Lys	Cys	Phe	Arg	Asn	Cys	Tyr	Ser	Leu	Asn	Tyr	
				395					400					405	
Ile	Ser	Asp	Val	Arg	Pro	Ala	Phe	Leu	Pro	Pro	Asp	Val	Tyr	Ala	
				410					415					420	
Asn	Asn	Ser	Tyr	Val	Asp	Leu	Asp	Val	His	Phe	Arg	Phe	Glu	Thr	
				425					430					435	
Ile	Met	Val	Tyr	Arg	Thr	Ser	Leu	Val	Phe	Gly	Trp	Val	Asp	Leu	
				440					445					450	
Met	Val	Ser	Phe	Gly	Gly	Ile	Ala	Gly	Leu	Phe	Leu	Gly	Cys	Ser	
				455					460					465	
Leu	Ile	Ser	Gly	Met	Glu	Leu	Ala	Tyr	Phe	Leu	Cys	Ile	Glu	Val	
				470					475					480	
Pro	Ala	Phe	Gly	Leu	Asp	Gly	Leu	Arg	Arg	Arg	Trp	Lys	Ala	Arg	
				485					490					495	
Arg	Gln	Met	Asp	Leu	Gly	Val	Thr	Val	Pro	Thr	Pro	Thr	Leu	Asn	
				500					505					510	
Phe	Gln	Gln	Thr	Thr	Pro	Ser	Gln	Leu	Met	Glu	Asn	Tyr	Ile	Met	
				515					520					525	
Gln	Leu	Lys	Ala	Glu	Lys	Ala	Gln	Gln	Gln	Lys	Ala	Asn	Phe	Gln	
				530					535					540	
Asn	Trp	His	Arg	Ile	Thr	Phe	Ala	Gln	Lys	His	Val	Ile	Gly	Lys	
				545					550					555	

<210> 28  
 <211> 2080  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1416107CB1

<400> 28  
 ggccggttcag ggcgcagagc tggccgatcg gcgttggccg ccgacatgac gcccagaggac 60  
 ccagaggaaa cccagccgct tctggggcct cctggcggca gcgcgccccg cggccgcccgc 120  
 gtcttccctc cgcccttcgc cgctgccctg ggccactca gcttcggett cgcgctcggc 180  
 tacagtcgcc cggccatccc tagcctgcag cgcgcgcgcg ccccgggccc gcgcttgac 240  
 gacgcgcgcg cctcctgggt cggggctgtc gtgaccctgg gtgcgcgcgc ggggggagtg 300



ctgggcccgt	ggctgggtga	ccgcgcgggg	cgcaagctga	gcctcttget	gtgctccgtg	360
cccttcgtgg	cgggctttgc	cgatcatcacc	gcggcccagg	acgtgtggat	gctgctgggg	420
ggccgcctcc	tcaccggcct	ggcctgcggg	gttgccctccc	tagtgggccc	ggtctacatc	480
tccgaaatcg	cctaccagc	agtccggggg	ttgctcggct	cctgtgtgca	gctaattggtc	540
gtcgtcggca	tcctcctggc	ctacctggca	ggctgggtgc	tggagtggcg	ctggctggct	600
gtgctgggct	gcgtgcccc	ctccctcatg	ctgcttctca	tgtgcttcat	gcccagagacc	660
ccgcgcttcc	tgctgactca	gcacaggcgc	caggaggcca	tggccgccct	gcggttccctg	720
tggggctccg	agcagggtcg	ggaagacccc	cccatcgggg	ctgagcagag	ctttcacctg	780
gcctgtctgc	ggcagcccgg	catctacaag	cccttcatca	tcggcgtctc	cctgatggcc	840
ttccagcagc	tgctgggggt	caacgcctgc	atgttctatg	cagagaccat	ctttgaagag	900
gccaaagtta	aggacagcag	cctggcctcg	gtcgtcgtgg	gtgtcatcca	ggtgctgttc	960
acagctgtgg	cggctctcat	catggacaga	gcagggcgga	ggctgctcct	ggtcttgtca	1020
ggtgtgggtca	tgggtgttcag	cacgagtgc	ttcggcgccct	acttcaagct	gaccagggt	1080
ggccctggca	actcctcgca	cgtggccatc	tcggcgccctg	tctctgcaca	gcctgttgat	1140
gccagcgtgg	ggctggcctg	gctggccgtg	ggcagcatgt	gcctcttcat	cgccggcttt	1200
gcggtgggct	ggggggccat	cccctggctc	ctcatgtcag	agatcttccc	tctgcatgtc	1260
aagggcgtgg	cgacaggcat	ctgcgtcctc	accaactggc	tcatggcctt	tctcgtgacc	1320
aaggagtcca	gcagcctcca	ggaggctcct	aggccctatg	gagccttctg	gcttgccctc	1380
gctttctgca	tcttcagtgt	ccctttcact	ttgttctgtg	tccctgaaac	taaaggaaag	1440
actctggaac	aaatcacagc	ccattttgag	gggcgatgac	agccactcac	taggggatgg	1500
agcaagcctg	tgactccaag	ctgggcccac	gccagagacc	cctgcctgcc	ccaggggagc	1560
cagaatccag	ccccttggag	ccttgggtctg	cagggtccct	ccttccctgtc	atgctccctc	1620
cagcccatga	cccggggcta	ggaggctcac	tgctcctgt	tccagctcct	gctgctgtc	1680
tgaggactca	ggaacacctt	cgagctttgc	agacctgcgg	tcagccctcc	atggccaaga	1740
ctaaagcagc	ggaagaggag	gtgggcctct	aggatctttg	tcttctggct	ggagggtgctt	1800
ttggagggtt	ggtgctgggc	attcagtcgc	tctcttcacg	cggtgcctt	atcggaagg	1860
aaatttgttt	gccaaataaa	gactgacaca	gaaaatcagg	tcagtgtctc	tgggctttgt	1920
gcaagctcag	ttgaaaagg	gtttattccc	atcatgccc	aggacaccct	gtggctttac	1980
ttgctcatgg	tcagccaagc	ttacccttca	cactgagaag	tcatttctgg	ctacttccct	2040
gggctcagtt	ccctgggtca	tcagccatca	aatcttgttg			2080

&lt;210&gt; 29

&lt;211&gt; 2128

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1682513CB1

&lt;400&gt; 29

ctggccctag	ggagctgccc	ctgtcgtctg	ctgcctgcac	caaccagccc	cacattgtca	60
actacctgac	ggagaacccc	cacaagaagg	cggacatgcg	gcgccaggac	tcgcgaggca	120
acacagtgtc	gcatgcgctg	gtggccattg	ctgacaacac	ccgtgagAAC	accaagtgtg	180
ttaccaagat	gtacgacctg	ctgctgtctc	agtgtgcccg	cctcttcccc	gacagcaacc	240
tggaggccgt	gctcaacaac	gacggcctct	cgccctcat	gatggctgcc	aagacgggca	300
agattgggaa	cgcgcaggag	atgctggctg	tggagcccat	caatgaactg	ctgcgggaca	360
agtggcgcaa	gttcggggcc	gtctcctctg	acatcaacgt	ggtctcctac	ctgtgtgcca	420
tggctcatctt	cactctcacc	gcctactacc	agccgctgga	gggcacaccg	ccgtaccctt	480
accgcaccac	ggtggactac	ctgcggctgg	ctggcagagg	cattacgctc	ttcactgggg	540
tcctgttctt	cttcaccaac	atcaaagact	tgttcatgaa	gaaatgccct	ggagtgaatt	600
ctctcttcat	tgatggctcc	ttccagctgc	tctacttcat	ctactctgtc	ctggtgatcg	660
tctcagcagc	cctctacctg	gcagggatcg	aggcctacct	ggcctgtatg	gtctttgccc	720
tggctcctggg	ctggatgaat	gccctttact	tcaccctgtg	gctgaagctg	acggggacct	780
atagcatcat	gatccagaag	attctcttca	aggacctttt	ccgattccctg	ctcgtctact	840
tgtcttctat	gatcggctac	gcttcagccc	tgttctccct	cctgaacccg	tgtgccaaca	900
tgaagggtgtg	caatggggag	cagaccaact	gcacagtgcc	cacttaccct	tcgtgccgtg	960
acagcgagac	cttcagcacc	ttctcctctg	acctgtttta	gctgaccatc	ggcatggggc	1020
acctggagat	gctgagcagc	accaagtacc	ccgtggctct	catcatcctg	ctggtgacct	1080
acatcatcct	caactttgtg	ctgctcctca	acatgctcat	tgcctcatg	ggcgagacag	1140
tgggccagggt	ctccaaggag	agcaagcaca	tctggaagct	gcagtggggc	accaccatcc	1200
tggacattga	ccgtctcttc	cccgtattcc	tgaggaaagt	cttccgctct	ggggagatgg	1260
tcaccgtggg	caagagctcg	gacggcactc	ctgaccgcag	gtgggtgctt	aggggtggatg	1320
aggtgaactg	gtctcactgg	aaccagaact	tgggcatcat	caacgaggac	ccgggcaaga	1380
atgagacctc	ccagtattat	ggcttctcgc	ataccgtggg	ccgcctccgc	agggatcgct	1440
ggtcctcggg	ggtacccccc	gtggtggaac	tgaacaagaa	ctcgaacccg	gacgaggtgg	1500
tgggtgcctct	ggacagcagc	gggaaccccc	gctgcgatgg	ccaccagcag	ggttaccctc	1560

gcaagtggag	gactgatgac	gccccgctct	agggactgca	gcccagcccc	agcttctctg	1620
cccactcatt	tctagtccag	ccgcatttca	gcagtgcctt	ctgggggtgtc	cccccacacc	1680
ctgcttttggc	cccagaggcg	agggaccagt	ggaggtgcca	gggaggcccc	aggaccctgt	1740
ggccccctgg	ctctgcctcc	ccaccctggg	gtggggggtc	ccggccacct	gtcttgtctc	1800
tatggagtca	cataagccaa	cgccagagcc	cctccacctc	aggccccagc	ccctgcctct	1860
ccattattta	tttgcctctg	tctcaggaag	cgacgtgacc	cctgccccag	ctggaacctg	1920
gcagaggcct	taggaccccg	ttccaagtgc	actgcccggc	caagccccag	cctcagcctg	1980
cgctgagct	gcatgcgcca	ccatttttgg	cagcgtggca	gctttgcaag	gggctggggc	2040
cctcggcggtg	gggccatgcc	ttctgtgtgt	tctgtagtgt	ctgggatttg	ccggtgctca	2100
ataaatgttt	attcattgaa	aaaaaaaa				2128

&lt;210&gt; 30

&lt;211&gt; 2825

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2446438CB1

&lt;400&gt; 30

cgttgtgcac	gtaattcggc	tgcacgtgtg	tccagatggg	cagtctctgg	tggttagcct	60
gtctcgacag	gggagagtta	agctcccggt	ctccaccgtg	ccggctgggc	aggtgggctg	120
aggggtgaccg	agagaccaga	acctgcttgc	tggagcttag	tgctcagagc	tggggaggga	180
ggttccgccc	ctctcttgc	gtcagcgccg	gcagcccctc	ccggcttcac	ttctccccc	240
agccccctgct	actgagaagc	tccgggatcc	cagcagccgc	cacgcccctg	cctcagcctg	300
cggggctcca	gtcaggccaa	caccgacgcg	cagctgggag	gaagacagga	cccttgacat	360
ctccatctgc	acagaggtcc	tggctggacc	gagcagcctc	ctcctcctag	gatgacctca	420
ccctccagct	ctccagtttt	caggttggag	acattagatg	caggccaaga	agatggctct	480
gaggcggaca	gaggaaagct	ggatttttgg	agcgggctgc	ctcccatgga	gtcacagttc	540
caggggcgagg	accggaaatt	cgccccctcag	ataagagtca	acctcaacta	ccgaaaggga	600
acaggtgcc	gtcagccgga	tccaaaccga	tttgaccgag	atcggtctct	caatgcggct	660
tcccgggggtg	tcccagagga	tctggctgga	cttccagagt	acctgagcaa	gaccagcaag	720
tacctcaccg	actcggaata	cacagagggc	tccacaggta	agacgtgcct	gatgaaggct	780
gtgctgaacc	ttaaggacgg	ggtcaatgcc	tgcattctgc	cactgctgca	gatcgaccgg	840
gactctggca	atcctcagcc	cctggtaaat	gcccagtgca	cagatgacta	ttaccgaggc	900
cacagcgctc	tgcacatcgc	cattgagaag	aggagtctgc	agtgtgtgaa	gctcctgggtg	960
gagaatgggg	ccaatgtgca	tgcccgggccc	tgcccgcgct	tcttccagaa	gggccaaggg	1020
acttgctttt	atttcgggtga	gctacccctc	tctttggccg	cttgccacca	gcagtgggat	1080
gtggtaagct	acctcctgga	gaaccacac	cagcccgcga	gcctgcaggc	cactgactcc	1140
cagggcaaca	cagtcctgca	tgccctagt	atgatctcgg	acaactcagc	tgagaacatt	1200
gcactgggtga	ccagcatgta	tgatgggctc	ctccaagctg	gggcccgcct	ctgcccctacc	1260
gtgcagcttg	aggacatccg	caacctgcag	gatctcacgc	ctctgaagct	ggccgcgaag	1320
gagggcaaga	tgcagatttt	caggcacatc	ctgcagcggg	agttttcagg	actgagccac	1380
ctttcccga	agttcacccga	gtggtgctat	gggcctgtcc	gggtgtcgct	gtatgacctg	1440
gcttctgtgg	acagctgtga	ggagaactca	gtgctggaga	tcattgcctt	tcattgcaag	1500
agcccgcacc	gacaccgaat	ggtcgthttg	gagcccctga	acaaactgct	gcaggcgaaa	1560
tgggatctgc	tcattcccga	gttcttctta	aacttctgt	gtaattctgat	ctacagtgtc	1620
atcttcaccg	ctgttgccca	ccatcagcct	acctgaaga	agcaggccgc	ccctcacctg	1680
aaagcggagg	ttggaaactc	catgctgctg	acgggccaca	tccttatcct	gctagggggg	1740
atctacctcc	tcgtgggcca	gctgtggtac	ttctggcgcc	gccacgtggt	catctggatc	1800
togttcatag	acagctactt	tgaaatcctc	ttctgttcc	aggccctgct	cacagtgggtg	1860
tcccagggtgc	tgtgtttcct	ggccatcgag	tggtacctgc	ccctgcttgt	gtctgcgctg	1920
gtgctgggct	ggctgaacct	gctttactat	acacgtggct	tccagcacac	aggcatctac	1980
agtgtcatga	tccagaagg	catcctgcgg	gacctgctgc	gcttccctct	gatctactta	2040
gtcttccttt	tcggcttcgc	tgtageccctg	gtgagcctga	gccaggaggc	ttggcgcccc	2100
gaagctccta	caggcccca	tgccacagag	tcagtgcagc	ccatggaggg	acaggaggac	2160
gagggcaacg	gggcccagta	cagggggtatc	ctgggaagcct	ccttggagct	cttcaaattc	2220
accatcgcca	tgggogagct	ggccttccag	gagcagctgc	acttccgcgg	catgggtgctg	2280
ctgctgctgc	tggcctacgt	gctgtccacc	tacatcctgc	tgctcaacat	gctcatcgcc	2340
ctcatgagcg	agaccgtcaa	cagtgtcgcc	actgacagct	ggagcatctg	gaagctgcag	2400
aaagccatct	ctgtcctgga	gatggagaat	ggctattggg	ggtgcaggaa	gaagcagcgg	2460
gcaggtgtga	tgtgacctg	tggcactaag	ccagatggca	gccccgatga	gcgctgggtc	2520
ttcagggtgg	aggaggtgaa	ctgggcttca	tgggagcaga	cgctgcctac	gctgtgtgag	2580
gacccgtcag	gggcagggtg	ccctcgaaat	ctcgagaacc	ctgtcctggc	ttccctctcc	2640
aaggaggatg	aggatgggtg	ctctgaggaa	aactatgtgc	ccgtccagct	cctccagctc	2700
aactgatggc	ccagatgcag	caggaggcca	gaggacagag	cagaggatct	ttccaaccac	2760

atctgctggc tctgggggtcc cagtgaattc tgggtggcaaa tatatatattt cactaaaaaa 2820  
aaaaa 2825

<210> 31  
<211> 1718  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2817822CB1

<400> 31  
gcctcggtgt tcccacctag gggcgggcag ccaggggcac ttccgctggc ccaagtgate 60  
tgcattgtggc agggctgcgc agtggagcgg ccagtgggca ggatgacgag ccagaccctt 120  
ctgccccagt ccccccgcc caggcgggca acgatgtcta ctgttggtga gctgaacgtc 180  
gggggtgagt tccacaccac caccctgggt accctgagga agtttccggg ctcaaagctg 240  
gcagagatgt tctctagctt agccaaggcc tccacggacg cggagggccg cttcttcac 300  
gaccgcccc gacctattt cagacccatc ctggactacc tgcgcactgg gcaagtgcc 360  
acacagcaca tccctgaagt gtaccgtgag gctcagttct acgaaatcaa gcctttgggtc 420  
aagctgctgg aggacatgcc acagatcttt ggtgagcagg tgtctcggaa gcagtttttg 480  
ctgcaagtgc cgggctacag cgagaacctg gagctcatgg tgcgcctggc acgtgcagaa 540  
gccataacag cacggaagtc cagcgtgctt gtgtgcctgg tggaaactga ggagcaggat 600  
gcatattatt cagaggctct gtgttttctg caggataaga agatgttcaa gtctgttgtc 660  
aagtttgggc cctggaaggc ggtcctagac aacagcgacc tcatgcactg cctggagatg 720  
gacattaagg cccaggggta caagggtatt tccaagttct acctgacgta cccaccacaa 780  
agaaacgaat tccattttta cattttattc ttcacctca cctgggtggtg atcctcagga 840  
gcagagactg ttatgaattc tggcgtggct tatgaaatta aaagttgcca tcaaagccat 900  
tttcttttaa tttcacaac atcaggcaat ttccagggtt ggtctagagt cttgccacta 960  
aatattgate actcgtttta ggactttcca ctccattgca actgatgcca ctatatttgc 1020  
ctagcaactt gcagctactt ccttttcaaa gctcatgta tctcccagac cttctctctg 1080  
aagtccaata acaagaccaa gtaagaatgt ttcaacaatg cgttggcaag agatgtgaga 1140  
tgacaacagg aacatacaag atactgtgaa tctagatggt ctgacctaaa gatgtagtct 1200  
acatagcccc agcttggggg ccaatccatc tgtccctggc atgtgccttc atgtagtagg 1260  
tgctttcctg atcccccttg cgagatgctg tgggtgctaa cacctcagag ctgtcctctt 1320  
ctctagagtg gaggttttca aagtgcacat tcagcattac ctgtgaactt gctggaaata 1380  
caaactctca gggcccacct cagacctact gaatcagaat ctctgggggt tggcacagca 1440  
ttctgattta ccaaaccctc caagtgatct tgatgtattc taattttgag accatctcta 1500  
gaaaagaatt gctacctctt gtatggaggt acaaaagact gacctcttac atcaaggaac 1560  
ttcctttccc agagctcctc atggaatcaa gctgaagtca gtcttctctt gagagcacat 1620  
tcttactcag ttttttctct ctgtcctacg ctgcttccct cactccctt ctcttaagag 1680  
cactccatca ataaaccact tgcacgagaa aaaaaaa 1718

<210> 32  
<211> 2000  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 4009329CB1

<400> 32  
gacgaatttg aaaccagggg gtgtcctggt tgaacttggt gccagataga gtaactcgga 60  
ctccagttgg aggggttcgg gagaaccata gaagaggaag ggccgtgtct tccgtggaca 120  
ggccaccgga gccgccagct gtttggaact gagctactgc agaaaggga gtggagagta 180  
agggccaggg cccgtggggg cagatggccg gcagaaggct gaatctgcgc tgggactga 240  
gtgtgctttg tgtgctgcta atggcgagga cagtgtctgg gactaggggc tctgtacag 300  
gagctcacat tagccccag tttccagctt caggtgtgaa ccagaccccc gtggtagact 360  
gccgcaagggt gtgtggcctg aatgtctctg accgctgtga ctctcatccg accaaccctg 420  
actgccacag tgatgggggg tacctggact acctggaagg catcttctgc cacttccctc 480  
ccagctctct cctctggct gtcactctct acgtttcctg gctgctctac ctgtttctga 540  
ttctgggagt caccgcagcc aagtttttct gcccacactt gtcggccatt tctaccacac 600  
tgaagctctc ccacaacgtg gcaggcgtca ccttctctggc atttgggaat ggtgcacctg 660  
acatcttcag tgccctgggt gccctctctg acccgcacac agccggcctg gcccttgggg 720  
cactgtttgg cgtggcgtg ctggttacca cagtgggtggc cggaggcatt accatcctac 780  
acccttcat ggtgcctcc aggccttct tcaggacat cgttttctac atggtggctg 840

tggttctgac	cttctctcatg	ctcttccgtg	gcagggtcac	cctggcatgg	gctctggggtt	900
acctgggctt	gtatgtgttc	tatgtgggtc	ctgtgattct	ctgcacctgg	atctaccaac	960
ggcaacggag	aggatctctg	ttctgcccc	tgccagttac	tccagagatc	ctctcagact	1020
ccgaggagga	ccgggtatct	tctaatacca	acagctatga	ctacgggtgat	gagtaccggc	1080
cgctgttctt	ctaccaggag	accacggctc	agatcctggg	ccgggccctc	aatccccctg	1140
attacatgaa	gtggagaagg	aaatcagcat	actggaaagc	cctcaagggtg	ttcaagctgc	1200
ctgtggagtt	cctgctgctc	ctcacagctc	ccgtcgtgga	ccgggacaag	gatgaccaga	1260
actggaaacg	gccccctcaac	tgtctgcctc	tggttatcag	cccctgtggtt	gtgggtcctga	1320
ccctgcagtc	ggggacctat	ggtgtctatg	agataggcgg	cctcgttccc	gtctgggtcg	1380
tggtgggtgat	cgcaggcaca	gccttggctt	cagtgcacct	ttttgccaca	tctgacagcc	1440
agccccccag	gcttcaactgg	ctctttgctt	tcttgggctt	tctgaccagc	gccccgtgga	1500
tcaacgcggc	cgcacacagag	gtgggtgaaca	tcttgcgggtc	cctgggtgtg	gtcttccggc	1560
tgagcaaac	tgtgctgggg	ctcacgctgc	tgccctggggg	gaacagcatt	ggagatgcct	1620
tctcggattt	cacactggct	cgcagggtg	acccacggat	ggcgttctcc	gcttgccttg	1680
gcgccatcat	cttcaacatc	ctcgtgggtg	tggggctggg	ctgcctgctc	cagatctccc	1740
gaagccacac	agaagtgaag	ctggagccag	acggactgct	ggtgtgggtc	ctggcaggcg	1800
ccctggggct	cagcctcgtc	ttctccctgg	tctcagctcc	attgcagtgc	ttccagctca	1860
gcagagtcta	tggcttctgc	ctgctcctct	tctacctgaa	cttcttctgc	gtggccctcc	1920
tcattgaatt	tggagtgaatt	cacctgaaa	gcattgtgact	gaagccgctt	agtgtctgtg	1980
cctcactgca	ggcaggagcc					2000

&lt;210&gt; 33

&lt;211&gt; 2216

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6618083CB1

&lt;400&gt; 33

gaaaactctt	cctgaaggag	atgcagagga	agattcgaac	tggaggaaaa	ccctaaaaata	60
aacaataaca	acaaaagtct	aaaacctgaa	aagtgaacca	tgaagctcag	taaaaaggac	120
cgaggagaag	atgaagaaag	tgattcagcg	aaaaagaaat	tggactgggtc	ctgctcgctc	180
ctcgtggcct	ccctcgcggg	cgccttcggc	tctcctctcc	tctacggcta	caacctgtcg	240
gtgggtgaatg	ccccaccccc	gtacatcaag	gcctttttaca	atgagtcattg	ggaaagaagg	300
catggacgtc	caatagaccc	agacactctg	actttgtctc	ggtctgtgac	tgtgtccata	360
ttcgccatcg	tgggacttgt	ggggacgtta	attgtgaaga	tgatttgaaa	ggttcttggg	420
aggaagcaca	ctttgtctgg	caataatggg	tttgcaattt	ctgctgcatt	gctgatggcc	480
tgctcgctcc	aggcaggagc	ctttgaaatg	ctcatcggtg	gacgcttcat	catgggcata	540
gatggaggcg	tgcgctcag	tgtgctcccc	atgtacctca	gtgagatctc	acccaaggag	600
atcctgtggt	ctctggggca	ggtgactgcc	atctttatct	gcattggcgt	gttccactggg	660
cagcttcttg	gcctgcccga	gctgctggga	aaggagagta	cctggccata	cctgttttgg	720
gtgattgtgg	tccctgcctg	tgtccagctg	ctgagccttc	cctttctccc	ggacagccca	780
cgctacctgc	tcttgagaga	gcacaacgag	gcaagagctg	tgaagccctt	ccaaacgttc	840
ttgggtaaa	cagacgtttc	ccaagaggta	gaggagggtcc	tggctgagag	ccacgtgcag	900
aggagcatcc	gccttgggtc	cgtgctggag	ctgctgagag	ctccctacgt	ccgctggcag	960
gtgggtcacg	tgattgtcac	catggcctgc	taccagctct	gtggcctcaa	tgcaatttgg	1020
ttctatacca	acagcatctt	tggaaaagct	gggatccctc	tggcaaagat	cccatacgtc	1080
accttgagta	cagggggcat	cgagactttg	gctgccgtct	tctctgggtt	ggtcattgag	1140
cacctgggac	ggagacccct	cctcattggg	ggcttttggg	tcatgggcct	cttcttttgg	1200
accctcacca	tcacgtgac	cctgcaggac	cacgcccctc	gggtccccta	cctgagtatc	1260
gtgggcattc	tggccatcat	cgcctctttc	tgcagtgggc	cagctgtttt	cccagaagaa	1320
acggtaaatg	tcagcattgt	atctgagtga	aaagttgacc	ttcttcccca	cccatgcaca	1380
caaacaagcc	agattggact	catctgcata	tctgcctgaa	gttcttttgc	aacaaaaaat	1440
cactaagctt	agccttctct	gttttttttt	tcctaagccc	tcccaagact	ttttgcaatg	1500
atcctgatte	tgttccaagt	gtttgcaact	gtggctttct	tttgactgta	gaacatgctg	1560
catttccagg	gctttaaatg	ctgggctccc	catcagtgtc	tatgggactc	cctggaggga	1620
aggccacctg	cacctcccaa	tcccagatca	cctgtcagcc	cctgccctcc	gcttctctca	1680
tccatcttca	acccctgtg	ttgaccagc	acctgggcct	tgctggctag	caatgacttt	1740
agccacaaga	tggaccaggg	tttagaagct	tcattttaa	tcacattgac	agtgtacagt	1800
ttaaagcttc	agggaacctt	cctgtctaa	aaaagctgcc	acttagacca	tgagaccatc	1860
ttgactcttc	ctaagtggac	agggaagag	aagtcgccag	gggagccacc	cgggaaagtg	1920
tggcaggaag	atgctcagag	ctgaatggca	gagagactca	tgggcctgct	ctccatgatt	1980
aaagaagagg	gatggatctc	ccaggagagg	gccaggaggc	cgcctgagge	agcttctgtg	2040
aggaacaggt	cgatgtaaga	agacttgaca	aggagttgaa	attaggtgaa	agcaaaagaa	2100
gaaaacaaga	gaggcagttt	cctgctgcac	attttatattg	tgtgcataac	cccaaggcag	2160

tggcagggaa gtctaataaa tgaggcaaaa taaaagagct tcacctttta aaaaaa 2216

<210> 34

<211> 1995

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7472002CB1

<400> 34

atgaccgaaa	aaaccaatgg	tgtgaagagc	tccccagcca	ataatcacaa	ccatcatgca	60
cctcctgcca	tcaaggccaa	tggcaaagat	gaccacagga	caagcagcag	gccacactct	120
gcagctgacg	atgacacctc	ctcagaactg	cagaggctgg	cagacgtgga	tgccccacag	180
caggggaagga	gtggcttccg	caggatagtt	cgctctgggg	ggatcatcag	agaatggggc	240
aacaagaatt	tccgagagga	ggaacctagg	cctgactcat	tcctcgagcg	ttttcgtggg	300
cctgaactcc	agactgtgac	cacacaggag	ggggatggca	aaggcgacaa	ggatggcgag	360
gacaaaggca	ccaagaagaa	atgtgaacta	tttgtcttgg	accagctggg	ggattgggtac	420
tactgctggc	tatttgtcat	tgccatgccc	gtcctttaca	actgggtgct	gctgggtggc	480
agagcctgct	tcagtgcact	acagaaaggc	tactacctgg	tgtggctggg	gctggattat	540
gtctcagatg	tgggtctacat	tgccggacct	ttcatccgat	tgccgacagg	tttcctggag	600
caggggctgc	tgggtcaaaga	taccaagaaa	ctgcgagaca	actacatcca	cacctgcag	660
ttcaagctgg	atgtggcttc	catcatcccc	actgacctga	tctattttgc	tgtggacatc	720
cacagccctg	aggtgcgctt	caaccgcctg	ctgcactttg	cccgcagtgt	tgagttcttt	780
gaccggacag	agacacgcac	caactaccct	aacatcttcc	gcacagcaa	ccttgtcttc	840
tacatcttgg	tcacatccca	ctggaatgcc	tgcactctat	atgccatctc	caaatccata	900
ggctttgggg	tcgacacctg	ggtttaccce	aacatcactg	accctgagta	tggctacctg	960
gctagggaat	acatctattg	cctttactgg	tccacactga	ctctcactac	cattggggag	1020
acaccacccc	ctgtaaagga	tgaggagtac	ctatttgtca	tctttgactt	cctgattggc	1080
gtcctcatct	ttgccaccat	cgtgggaaat	gtgggctcca	tgatctccaa	catgaatgcc	1140
accggggcag	agttccaggc	taagatcgat	gccgtgaaac	actacatgca	gttccgaaag	1200
gtcagcaagg	ggatggaagc	caaggctcatt	aggtgggttg	actacttgtg	gaccaataag	1260
aagacagtgg	atgagcgaga	aattctcaag	aatctgccag	ccaagctcag	ggctgagata	1320
gccatcaatg	tccacttgte	cacactcaag	aaagtgcgca	tcttccatga	ttgtgaggct	1380
ggcctgctgg	tagagctggg	actgaaactc	cgctctcagg	tcttcagtc	tggggattac	1440
atltgcccga	aaggggacat	cggcaaggag	atgtacatca	ttaaggaggg	caaactggca	1500
gtggtggctg	atgatgtgtg	gactcagtat	gctctgctgt	cggctggaag	ctgctttggc	1560
gagatcagta	tccttaacat	taagggcagt	aaaatgggca	atcgacgcac	agctaataatc	1620
cgcagcctgg	gctactcaga	tctcttctgc	ttgtccaaag	atgatcttat	ggaagctgtg	1680
actgagtacc	ctgatgccaa	gaaagtccta	gaagagaggg	gtcgggagat	cctcatgaag	1740
gagggactgc	tggatgagaa	cgaagtggca	accagcatgg	aggtcgacgt	gcaggagaag	1800
ctagggcagc	tggagaccaa	catggaaacc	ttgtacactc	gctttggccg	cctgctggct	1860
gagtacacgg	gggccccagca	gaagctcaag	cagcgcacat	cagttctgga	aaccaagatg	1920
aaacagaaca	atgaagatga	ctacctgtct	gatgggatga	acagccctga	gctggctgct	1980
gctgacgagc	cataa					1995

<210> 35

<211> 988

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1812692CB1

<400> 35

cttgggtgaa	agaaaatcct	gcttgacaaa	aaccgtcact	taggaaaaga	tgtcctttcg	60
ggcagccagg	ctcagcatga	ggaacagaag	gaatgacact	ctggacagca	ccgggaccct	120
gtactccagc	gcgtctcgga	gcacagactt	gtcttacagt	gaaagcgact	tgggtgaattt	180
tattcaagca	aattttaaga	aacgagaatg	tgtcttcttt	accaaagatt	ccaaggccac	240
ggagaatgtg	tccaagtgtg	gctatgccc	gagccagcac	atggaaggca	cccagatcaa	300
ccaaagtgag	aaatggaaact	acaagaaaca	caccaaggaa	tttctaccg	acgccttttg	360
ggatattcag	tttgagacac	tggggaagaa	agggaagtat	atacgtctgt	cctgcgacac	420
ggacgcggaa	atcctttacg	agctgctgac	ccagcactgg	cacctgaaaa	cacccaacct	480
ggctcattct	gtgaccgggg	gcgccaaagaa	cttcgccttg	aagccgcgca	tgcgcaagat	540
cttcagccgg	ctcatctaca	tcgcgcagtc	caaaggtgct	tggattctca	cgggaggcac	600

ccattatggc	ctgatgaagt	acatcgggga	ggtggtgaga	gataacacca	tcagcaggag	660
ttcagaggag	aatattgttg	ccattggcat	agcagcttgg	ggcatggtct	ccaaccggga	720
caccctcatc	aggaattgcg	atgctgaggt	accggtggga	caggaggagg	tctgctaggt	780
cacatggaag	aaagaccatg	gcatgggcct	gtggcctgaa	ccctggggct	ctgtgatgga	840
gccagccaga	tcattgggaa	gtctgccttt	caaggagtgc	ctttggggacc	ttaaagggaat	900
tgaaaacaag	gatgacgtac	ctaattaact	gctgggaaag	agttaacaat	gaatgttttg	960
ttcattaata	tgtgtttctca	gcaatctc				988

&lt;210&gt; 36

&lt;211&gt; 3179

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3232992CB1

&lt;400&gt; 36

gcggagcggc	ggcgccggcg	ccggggggcg	cagcgagggg	ctggcggtag	cggttgctgc	60
ggggcgcggg	gcgcgggcg	cgctggagtc	tcggcccgcg	gcgatgaggt	gcagacgctg	120
tcgggcagcg	taaggcgggc	cccgaaccga	ccccccggca	cccccgccac	ccccggctgc	180
gcagctactg	caaagggggc	ccggcgctca	gcagcccaaa	ccggccagct	tggggccggcg	240
gcggggcgga	agcgccggca	atcctcagct	tgggcaacgt	gctcaactac	ctggacaggt	300
acaccgtggc	aggcgctcct	ctggacatcc	agcagcactt	tgggggtcaag	gaccgaggcg	360
ccggcctgct	gcagtcagtg	ttcatctgta	gcttcatggt	ggctgcccc	atcttcggct	420
acctgggcga	ccgcttcaac	aggaaggtga	ttctcagctg	cggcattttc	ttctggtcgg	480
ccgtcacctt	ctccagctcc	ttcattcccc	agcagtaact	ctggctgctg	gtcctgtccc	540
gggggctggg	gggcacccgg	gaggccagct	actccaccat	cgccccact	atcattggcg	600
acctcttcac	caagaacacg	cgtacgctca	tgtgtccgt	cttctacttc	gccatccac	660
tgggcagtg	cctgggctac	attactggct	ccagcgtgaa	gcaggcagcc	ggagactggc	720
actgggcatt	gcgggtgtcc	cctgtcctgg	gcatgatcac	aggaacactc	atcctcattc	780
tggctccagc	cactaaaagg	ggtcatgccc	accagctcgg	ggaccagctc	aaggcccgga	840
cctcatggct	ccgagatatg	aaggccctga	ttcgaaaccg	cagctacgtc	ttctcctccc	900
tggccacgtc	ggctgtctcc	ttcgccacgg	gggcccctgg	catgtggatc	ccgctctacc	960
tgcaccgcgc	ccaagttgtg	cagaagacag	cagagacgtg	caacagcccc	ccctgtgggg	1020
ccaaggacag	cctcatcttt	ggggccatca	cctgctttac	gggatttctg	ggcggtgtca	1080
cgggggcagg	agccacgcgc	tgggtgcgcg	tgaagacca	gcgggcccag	ccactgggtg	1140
gtgcccgtgg	catgctgggg	tctgccatct	tcactctgct	gatcttcgtg	gctgccaaag	1200
gcagcatcgt	aggagccctat	atctgtatct	tcgtcgggga	gacgctgctg	ttttctaaact	1260
gggccatcac	tgcagacatc	ctcatgtacg	tggtcatccc	cacgcggcgc	gccactgccg	1320
tggccttgca	gagcttcacc	tcccacctgc	tgggggacgc	cgggagcccc	tacctcattg	1380
gctttatctc	agacctgatc	cgccagagca	ccaggactc	cccgtctctg	gagttcctga	1440
gcctgggcta	cgcgctcatg	ctctgccctt	tcgtcgtggt	cctggggcgg	atgttctctc	1500
tcgccactgc	gctcttcttc	gtcagcgacc	gcgccagggg	tgagcagcag	gtgaaccagc	1560
tggcgatgcc	gcccgcctct	gtgaaagtct	gaggtggtgc	cattggggaca	atgaagaacc	1620
cacactccca	cctcgtctgg	gaggtgtcct	acagcgtccg	ggaccggctg	ggctgcccc	1680
aagctttctg	tgtgatccac	ggctaggcac	ccacctctc	tggcccaggc	ctgctgagtg	1740
gccctggcat	caagaggagg	ctgtgtcctc	agttaccctg	gaaggatgtg	tgtgttggag	1800
ccacacgggt	ggacagggtt	ccagccctag	gtttggggcc	cagggcccc	ggggccaagg	1860
aagaagacag	ccccaaagtg	gtgtccgggg	agagcctggc	ctgccaccag	cttatgtgat	1920
cttgggcaag	tccctgccct	ccctggaacg	aaggggccagg	gggctggact	ttcccacaca	1980
acttgcctgg	caaagcacga	tctgcagctt	tgaagactca	acagaccctg	gaccatacgg	2040
agagcaggtg	gcccaggcct	cagggcgcca	gtcccggctt	tgaggctcac	gcgagggcct	2100
ggtatgcagg	gacctgtct	cagctgggce	tcggaccttg	gggatatttg	acgcaacctg	2160
gcaaataga	ctgggcgccc	aagtctcttg	gtactccctg	gaggacactg	tctcactgtc	2220
tcgggttggc	tcccagcctg	gaggtcccag	atggggactg	ttctgacaag	ctggcatcac	2280
caggggtgaa	ggccctgggt	gcagctgtac	accacctgtg	ccccaggct	caaggctctc	2340
ggcaggtgca	caccagccca	actctgcagg	gcttctctcc	ctgccaccac	cccccaagcc	2400
aggaccccac	tccttccccg	aggctgagct	gagccttttc	caggggcagg	gcccaggaga	2460
ccattcccag	aatccatggg	gcagtagcca	gggtccgggc	tgctggaggga	agcagctatc	2520
cacaaagctt	cctgccccag	agctgaggct	gagggccccg	gagaggcggc	ccctacccaa	2580
acactgctg	ctggcattcc	accaagtgc	cccaggggcc	aggccttcga	tcacccacct	2640
cccatccatg	cacacaccag	gatgcagctg	ccaacttcac	accagcccca	acccgctttg	2700
ggggagctta	gccccctg	tcacccactg	cctgcacttc	tgctgcaatc	aagggtggtc	2760
tgggtgcggg	gtggggtggg	gggtgaggcc	ttgtggccaa	tgggggaccc	cccaagagcc	2820
agcttggaca	atgctcttct	tgccccctag	ttactggctg	gctgtggctt	cagtgtgtgt	2880
taagcaggtg	gaatactcac	ccaccaagct	ctgggggtacc	ccgagggcct	gacaagagga	2940

tgggggtgggg	gtggcaccct	ccaaagacca	gcctccaccc	ccactccagc	ctcagcgggg	3000
ccccagcgat	gttttcttgt	tgtacaagaa	ccagggtccga	gtgttgccctc	ctcttccttc	3060
cggagccaa	actgctcctt	tatttttttag	agctgctgat	tgtgaatctc	agagtcttaa	3120
gagagaagcc	aaatatattc	ctcttgtaaa	tgaagaaata	aacctattta	aatcacaaa	3179

&lt;210&gt; 37

&lt;211&gt; 1986

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3358383CB1

&lt;400&gt; 37

ggagtatctg	agcaaattat	ttcttacctg	acttttagaga	aaacggctac	ctatctgacc	60
ccaaaacgac	ttgaggaaac	tgtttccacg	gtcctgctgc	agggggggaag	cacagtcgtc	120
aagaagagag	tggggctcagg	atcaaaacac	atttagtggtg	acttagggaa	agaaaacatt	180
ttccctcttt	gaacctctct	ggatacagtc	attttgccctc	tacttgagga	tcaactgttc	240
aacctcaatg	gcctttcagg	acctcctggg	tcacgctggg	gacctgtgga	gattccagat	300
ccttcagact	gtttttctct	caatctttgc	tggtgctaca	taccttcatt	ttatgctgga	360
gaacttcact	gcattcatac	ctggccatcg	ctgctgggtc	cacatcctgg	acaatgacac	420
tgtctctgac	aatgacactg	gggcccctcag	ccaagatgca	ctcttgagaa	tctccatccc	480
actggactca	aacatgaggc	cagagaagtg	tcgtcgcttt	gttcacctc	agtggcagct	540
ccttcacctg	aatgggacct	tccccaacac	aagtgacgca	gacatggagc	cctgtgtgga	600
tggtctgggtg	tatgacagaa	tctccttctc	atccaccatc	gtgactgagt	gggatctggg	660
atgtgactct	caatcactga	cttcagtggc	taaatttgta	ttcatggctg	gaatgatggg	720
gggaggcatc	ctaggcggtc	atctatcaga	caggtttggg	agaagggttcg	tgctcagatg	780
gtgttacctc	caggttgcca	ttgttggcac	ctgtgcagcc	ttggctccca	ccttcctcat	840
ttactgctca	ctacgcttct	tgtctgggat	tgctgcaatg	agcctcataa	caaatactat	900
tatgttaata	gccgagtggg	caacacacag	attccaggcc	atgggaatta	cattgggaat	960
gtgcccttct	ggtattgcat	ttatgacctc	ggcaggcctg	gcttttgcca	ttcgagactg	1020
gcatacctc	cagctgggtg	tgtctgtacc	atactttgtg	atctttctga	cctcaagttg	1080
gctgctagag	tctgctcggt	ggctcattat	caacaataaa	ccagaggaag	gcttaaagga	1140
acttagaaaa	gctgcacaca	ggagtgggaat	gaagaatgcc	agagacaccc	taaccctgga	1200
gattttgaaa	tccaccatga	aaaaagaact	ggaggcagca	caaaaaaaaa	aaccttctct	1260
gtgtgaaatg	ctccacatgc	ccaacatatg	taaaaggatc	tccctcctgt	cctttacgag	1320
atttgcaaac	tttatggcct	attttggcct	taatctccat	gtccagcatc	tgggggaaca	1380
tgttttcctg	ttgcagactc	tctttgggtc	agtcactctc	ctggccaact	gtgttgccac	1440
ttgggcaactg	aaatacatga	cccgtcgagc	aagccagatg	cgtctcatgt	acctactggc	1500
aatctgcttt	atggccatca	tatttgtgcc	acaagaaatg	cagacgctgc	gtgagggttt	1560
ggcaacactg	ggcttaggag	cgtcggctct	gaccaatacc	cttgcttttg	cccattggaa	1620
tgaagtaatt	cccaccataa	tcaggggcaag	agctatgggg	atcaatgcaa	cctttgctaa	1680
tatagcagga	gccctggctc	ccctcatgat	gacctaagt	gtgtattctc	cacctctgcc	1740
ctggatcatc	tatggagtct	tccccttcat	ctctggcttt	gctttcctcc	tccttctctga	1800
aaccaggaac	aagcctctgt	ttgacaccat	ccaggatgag	aaaaatgaga	gaaaagaccc	1860
cagagaacca	aagcaagagg	atccgagagt	ggaagtgcag	cagttttaag	gaattccagg	1920
agctgactgc	cgatcaatga	gccagatgaa	gggaacaatc	aggactattc	ctagacacta	1980
gcaaat						1986

&lt;210&gt; 38

&lt;211&gt; 3294

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4250091CB1

&lt;400&gt; 38

tgtaagacag	gaaagggatc	tattttgatgt	ctatcttcag	atatattggc	agttttcctt	60
aagctattta	gttctctcat	tgttgctttt	tcatttttgta	tactgcaagt	tcccaggcaa	120
ctcgaatttg	caaacacagc	catggataca	ctattttacct	tacagttagt	tcttggggaat	180
ctaagtctgg	tttttgttat	tcttccctcc	cctccactgc	ataatcatgt	ataactagca	240
acattttatg	ttataggttg	atttcctaag	tgtggctgat	ggtagcctct	agttttgaagt	300
gagggagaa	tgagttagtca	ggaactggtc	actttgaatg	tgaggaggaa	gatattcacg	360
acaaggtttt	ctacgataaa	gcagtttctt	gcttctcggt	tggcacgcac	gttagatggc	420

```

agagaccaag aattcaagat gggtgggtggc cagatTTTTT tagacagaga tgggtgatttg 480
tttagtttca tcttagattt tttgagaact caccagcttt tattaccac tgaattttca 540
gactatctta ggcttcagag agaggctctt ttctatgaac ttcgttctct agttgatctc 600
ttaaacccat acctgctaca gccaagacct gctcttgagg aggtacattt cctaagccgg 660
aacactcaag cttttttcag gggttttggc tcttgacgca aaacaattga gatgctaaca 720
gggaggatta cagtgtttac agaacaacct tcggcgccga cctggaatgg taacttttct 780
cctcctcaga tgaccttact tccactgcct ccacaaagac cttcttacca tgacctgggt 840
ttccagtgtg gttctgacag cactactgat aaccaaactg gagtcaggta ttttgtactt 900
tgcagtattt ctcttgata ccagtttggt atgttttctc taaaaacttg aagttcctca 960
ggcctgtaac ttctggaaaa gatgattatt caaaataatg ttttggggta accagtggag 1020
ttgggtagaa tgaccaaata attattttcc aaactgggat acttttttaga gtgaaagggg 1080
ctattattag gtgggacaaa aggaataaat gaagactgcc cagaaaaaac tgagactatg 1140
gacattcaaa tcatgggaga aaataatttt gtagattatg ttccattgct aatgaatttg 1200
acttagaaaa gaattgcctt atttttaaga gattgtttca gtggttcaca taaaggctcg 1260
ctcactgggt tctcttgagt tccttacaca ctatataagt tgttctttca gttttatgat 1320
tcaactactg tttttccttc agctgacttt atttttaaac acccttaaag acagatatat 1380
ctcatggcaa atttgggtat ctgttacagc cttggctctt aaacaactca aaatatggg 1440
ataggctgtc agtagttaa ggatagttgc tcctgagtca attcttctact tactccctct 1500
gttggtcttg gctggatcct aacgctgatt tccactctgc tgtcacaaac atttttcccc 1560
ccgtaaaatg tcttaatgct gtcctaccat tattttacca actgtgaaag ctggctttta 1620
tttttaggag gaaaagaaaa gcctgcatgt gttctttatt ggtatcattt aaaatatact 1680
tttttttttt ttttggtaaa ggtaggcgta ttttaagata ttttcttaac ttgagcagta 1740
gccaacagga aggataccag tgtctctctc tcctagcgac acactccttg gtccttgctta 1800
ccaactggag gacactagg agataaacog agtatgacaa ttcttaattg ttacattttt 1860
ataacttctt gtccttcaaa agagtgtgaa atgtcatttt gggaaaagag agccagtcaa 1920
gctagtaggc tgattgtgaa gaaaatctaa taccttatct ttatctcaaa cctctgtaca 1980
actttatttt cattgatggg atactttaac aaaaatgaaa ttttttttgg tttttaaaat 2040
atgagtattt atgacctctt tggggatcat gtttcaaaaa gtcagaaaac tagagacaaa 2100
actgtcattg atttttaaga agaaacacac taggtcaaaa gaagatgtcc tggaaatacg 2160
aagtactctt taaaaaccat gcatttggag aaagtaattg tttccttgaa aaacatgatt 2220
aaaaactaaa actgggatgt tcctgtgtgt acacagtgcc aaatggtttt ccctttttat 2280
gttggtgttt agaaacagca cgaaagtttt ttccatttta aagtgagaaa acattatatt 2340
tagacttcca taattccaaa atcagaagct atttttaaaa ttagcatttt ctgcatcac 2400
caaatgggat tcaattgttt gaagctcaaa atttttacca ttccataaat gtttgtgaat 2460
ttttagacag tgccaattta aaagtagaga tagccaatct gaatacgggt gaattatggg 2520
gatctctggt gattgggatg aaaactctgg ccttaaaagg tccactttta gtatataatt 2580
gcctaattag caatcatttt tattttttgc tcaactccctg gtctgaatct atctgtctat 2640
tcagatattt tttggtaggt ttggaaaatg gagaaagtga cctaattggg gcctaattgt 2700
ctggtgtatc attcacttta ttcagtttgt totatcaata tgatttacc ccaaggtta 2760
acctagcagg ttgctcagtt attatctctc aaggctcacag tactagaaat acttggcttg 2820
catctttcag atgccattca tgttatcaag ctcaaattat agttggtcac aggttctaa 2880
agtctttatt tgacttctcc tttttgaagt ggctcaaatg gaaaagtgtg gttgctttta 2940
aatgttaaaa ataagtttaa acttttatatt tccattgggt tccccctatt ttgtccttct 3000
tttgtgtgct tgaaatattt tatttttcag tttgtcctca tagggaatca agtatatttag 3060
ctaggtgatg tcttgcaagt acgttccact ttgttacaat ctactatctg tatatactat 3120
ttgtatctta attcttttat gagatgttct gtaacatttt tctcactttg acaaatgttt 3180
ttagactgta ctggcgcttg ggggtaagtg gaatgatttg ctaatatgta 3240
gaatctgttg tatcaaacat aataaacttt ttttgagatg tgaaaaaaa aaaa 3294

```

&lt;210&gt; 39

&lt;211&gt; 2043

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70064803CB1

&lt;400&gt; 39

```

gcaacatggc ggctgccgtg gtgcagcgcc cgggctgagc gacagcaagt gcagcgggct 60
cctaccccggt gtgaggggtg gcctccgcgt gggactgtgc cctcttcage ccgtcctgt 120
ccccgacatc acgtgtattc cgcacgtccc cctccgcgtg tgtgtctact gagacggga 180
ggcgtgacag ggcccggttc ccttctcagt ggtgctctgt gcttcagggc aagctccccg 240
tctccggggc cacttccctc gcctgtgttc ggtccatcct cctttctcca gcctcctccc 300
ctcgcaggty ggatcgctcg tgggaccgga gcgcggggcg gcgcggcccc cggggaccat 360
ggccgggtcc gacaccgcgc ccttctcag ccaggcggat gaccgggacg acgggaccagt 420
gcctggcacc ccggggttgc cagggtccac ggggaacccg aagtcggagg agcccgagg 480

```



```

ccccgaccag gagggggtgc agcgcatcac cggcctgtct cccggccggt cggctctcat 540
agtggcggtg ctgtgctaca tcaatctcct gaactacatg gaccgcttca ccgtggctgg 600
cgtccttccc gacatcgagc agttcttcaa catcggggac agtagctctg ggctcatcca 660
gaccgtgttc atctccagtt acatgggtgtt ggcacctgtg tttggctacc tgggtgacag 720
gtacaatcgg aagtatctca tgtgcggggg cattgccttc tgggtccctgg tgacactggg 780
gtcatecttc atccccggag agcattttctg gctgctcctc ctgaccgggg gcctgggtgg 840
ggtcggggag gccagttatt ccaccatcgc gccactctc attgccgacc tctttgtggc 900
cgaccagcgg agccggatgc tcagcatctt ctactttgcc attccgggtg gcagtggctc 960
gggtacaccc gtcaggtcca aagtgaagga tatggctgga gactggcact gggctctgag 1020
ggtagacacc ggtctaggag tgggtggcgt tctgctgctg ttccctggtag tgcgggagcc 1080
gccaaagggga gccgtggagc gccactcaga tttgccaccc ctgaacccca cctcgtgggt 1140
ggcagatctg agggctctgg caagaaatct catctttgga ctcatacact gcctgaccgg 1200
agtcttgggt gtgggcctgg gtgtggagat cagccgcggg ctccgccact ccaacccccg 1260
ggctgatccc ctggtctgtg ccactggcct cctgggctct gcaccttcc tcttctgtc 1320
ccttgctgc gccctgtgta gcatcgtggc cacttatatt ttcacttca ttggagagac 1380
cctcctgtcc atgaactggg ccactcgtggc cgacattctg ctgtacgtgg tgatccctac 1440
ccgacgtccc accgccgagg ccttccagat cgtgctgtcc cactgctgg gtgatgctgg 1500
gagccctac ctcaattggc tgatctctga ccgcctgcgc cggaaactggc cccctcctt 1560
ctgttccgag ttcgggctc tgcagttctc gctcatgctc tgcgcgtttg tgcggcact 1620
gggcggcgca gcccttctgg gcaccgccat cttcattgag gccgaccgcc ggccggcaca 1680
gctgcacgtg cagggcctgc tgcacgaagc agggctccaca gacgaccgga ttgtgggtgc 1740
ccagcggggc cgtccacccc gcgtgcccgt ggccagtgtg ctcatctgag aggcctgccg 1800
tcacctacct gcacatctgc cacagctggc cctgggcccc cccacgaag ggcctgggac 1860
taaccccttg cctggcccca gcttccagag cgtgtgtcca gctcccagac 1920
actacatggg tagctcaggg gaggaggtgg ggtccagga gggggatccc tctccacagg 1980
ggcagcccca agggctcggg gctatttgta acggaataaa atttgtagcc agacaaaaaa 2040
aaa

```

&lt;210&gt; 40

&lt;211&gt; 1915

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70356768CB1

&lt;400&gt; 40

```

caccactggg cgtgcgcgc tgccttccc tccgcgcaca ggtgcgggc tcaccgcttg 60
ctaattggcag ccgggggtct cctgggacag caagacctcc gctcaggccc ctctttcgaa 120
tgctccacgc cctcctgcga tctagaatga ttcagggcag gatcctgctc ctgaccatct 180
gcgtgcgcgg cattggtggg acttttcagt ttggctataa cctctctatc atcaatgccc 240
cgaccttgca cattcaggaa ttcaccaatc agacatggca ggcgcgtact ggagagccac 300
tgcccgatac cctagtctctg cttatgtggt cctcctcgtg gtctctgtat cccctgggag 360
gcctcttttg agcactgctt gcaggctcct tggccatcac gctgggaagg aagaagtccc 420
tcttggtgaa taacatcttt gtggtgtcag cagcaatcct gtttggtatc agccgcaaag 480
caggctcctt tgagatgata atgctgggaa gactgctcgt gggagtcaat gcagggtgta 540
gcatgaacat ccagccatg tacctggggg agagccccc taaggagctc cgaggagctg 600
tggccatgag ctgagccatc tttacggctc tggggatcgt gatgggacag gtggctggac 660
tcaggagact cctaggtggc cctcaggcct ggcccctgct gctggccagc tgccctgtgc 720
ccggggcgct ccagctcgcc tccctgcctc tgctccctga aagcccgccg tacctcctca 780
ttgactgtgg agacaccgag gctgcctgg cagagacggg tctcgccttg tccaggtgg 840
agtgcgtggt ctgttcatag gcatgacccc attgttgata agcacggaag ttttcttctt 900
ttttgttttt gttttttttg ttttgttttg gacggggtct cactctgtcg cccaggctgg 960
agtgggtgta tctcggtctg ctgcagcctc cacctcccgg gcccaatcgg ttctcccgcc 1020
tcagcctcct ggggtggctgg gactgctggc ccgtgccacc acgcttggct aatttttttt 1080
tattattgta ttttttgtaa agatggagtt tcacctcttt gctgggcag gtctcaaact 1140
cctgagatca aatgatcctc ccccttggc ctcccaaagt gcgtggatta taggcagtag 1200
ccattgtatc tggctagcat gggagttttg aactgtccca tttccaacct gggccagtgc 1260
attcctcctt aggcagcctg gtggtccctg ctccctgggag gtcactatat tgatgctgaa 1320
cttagtgtag acacctgata tgcctagcgt actgcaaccc agagctcctg ggcccaggcg 1380
atcctctgtg ctgagctcct tgagtagctg ggcacttagg cacacaccac tatgcgtggc 1440
tctccatgct ccttgggtct accctctgag atgtttttcc ttttctttca ccttccctga 1500
ttccttctga agagggcggt gcacaatgtg ctgcttttga tggttgagca aatttctcag 1560
cctccttctt gcctatagag agttggggca ggctggggcg cagctcacgc ctgtaatccc 1620
agggaggctg aggcgggcag atcacgaggt caggacatca agaccggcct ggccgacatg 1680
gtgggacccc atctctacta acaatacaaa aattggctgg gtatgggtgg acgtgcctgt 1740

```

ggtccccggt	gctggggagg	ctgaggcggg	agagttgctt	gggccccgga	ggcggagggtt	1800
gcagtggcgg	gagaattgct	tggggcccgg	gaggcggagg	ttgcggtgag	ccgagattgt	1860
gccagtgcac	actgcactcc	agcctggtga	cagagtgaga	ctccgtcttc	aaaaa	1915

<210> 41  
 <211> 1809  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5674114CB1

<400> 41

atgggcctgg	ccagggccct	acgccgcctc	agcggcgccc	tggattcggg	agacagccgg	60
gcgggcgatg	aagaggaggc	cgggcccggg	ttgtgccgca	acgggtgggc	gccggcaccg	120
gtgcagtcac	ccgtgggccc	gcgcccgggt	cgcttcgtca	agaaaagacg	gcactgcaac	180
gtgcgtttcg	taaacctggg	tggccagggc	gcgcgctacc	tgagcgacct	gttcaccaca	240
tgcgtggacg	tgcgctggcg	ctggatgtgc	ctgctcttct	cctgctcctt	cctcgccctcc	300
tggctgctct	tggcctgggc	cttctggctc	attgectcgc	tgacaggcga	cctggccgcc	360
ccgccaccgc	ccgcgcctcg	cttctcacac	gtggccagct	tccctggccg	cttccctcttc	420
gcgctggaga	cgcagacgtc	catcggctac	ggcgtgcgca	gcgtcaccga	ggagtggccc	480
gccgctgtgg	ccgcctgtgt	gctgcagtg	attgcccggc	gcgtgctcga	cgccttcgtc	540
gtgggtgtcg	tcattggccaa	gatggccaaa	cccaagaagc	gcaacgagac	gctgggtcttc	600
agcgagaaac	ccgtcgtggc	gctgcgcgac	caccgcctct	gcctcatgtg	gcgcgtcggc	660
aacctgcgcc	gcagccacct	ggtcgaggcc	cacgtgcgtg	cccagctgct	gcagccccgt	720
gtgacccacg	agggtgagta	catcccgtcg	gaccaccagg	atgtggatgt	gggctttgat	780
ggaggcaccc	atcgatatct	cctcgtgtcc	cccaccacca	tcgtccatga	gacgcactct	840
gccagtcctc	tgtatgagct	aggacgtgcc	gagctggcca	gggctgactt	tgagctgggtg	900
gtcattctcg	aggggatggg	tgaggccaca	gccatgacca	cacagtgtcg	ctcgtcctac	960
ctccctgggtg	aactgctctg	gggccatcgt	tttgagccag	ttctcttcca	gcgtgggtcc	1020
cagtatgagg	tcgactatcg	ccacttccat	cgcacttatg	aggtcccagg	gacaccggtc	1080
tgcagtgtca	aggagctgga	tgaacgggca	gagcaggctt	cccacagcct	caagtctagt	1140
ttccccggct	ctctgactgc	attttgttat	gagaatgaac	ttgctctgag	ctgctgccag	1200
gaggaagatg	aggacgatga	gactgaggaa	gggaatgggg	tggaaacaga	agatggggct	1260
gctagccccc	gagttctcac	accaaccctg	gcgctgaccc	tgccctccatg	atgcaaactg	1320
atgtccccct	ccccgtgtat	gcccccttcc	ccaaggtagc	aagatggagg	gatggggctc	1380
tctcctggga	tgggggcagg	tggtcctgaa	taccgacagg	cctgctgggt	aaatgactag	1440
gtggtaaggt	tctgccatgc	ctggtgaccc	accatggaca	tactggacct	taattcctct	1500
gcttctgtgc	tccctcctga	gaacccttta	tgagcctgat	tccctcagtc	caccagaatt	1560
ctggatcacc	caagaggaaa	agactggcag	ttctagatct	ctctatatgg	ggagacctgg	1620
attgttgacc	agggtgagaa	gccaatggta	tagactgcct	ctggggaagc	aagttggcag	1680
ttcttgaaac	gcatacagata	tcaagagttt	gtaggtctgg	attcacctaa	gattcaaggg	1740
agtgttgctt	ctcaactcag	ccaactgagt	agcaaatact	ttgttctaga	ccacctaaagg	1800
aggggaaggt						1809

<210> 42  
 <211> 1730  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1254635CB1

<400> 42

ctttggccta	ttataccatg	gatgctaaaa	atgggttctaa	ctgaaaaccc	aaaccaagaa	60
atagcaacaa	gtctagaatt	cttactacta	caaaactcac	ctggatccct	aagggcacag	120
caaagaatga	gctattacgg	cagcagctat	catattatca	atgcggacgc	aaaataccca	180
ggctacccgc	cagagcacat	tatagctgag	aagagaagag	caagaagacg	attacttcac	240
aaagatggca	gctgtaattg	ctacttcaag	cacatttttg	gagaatgggg	aagctatgtg	300
gttgacatct	tcaccactct	tgtggacacc	aagtggcgcc	atatgtttgt	gatattttct	360
ttatcttata	ttctctcgtg	gttgatattt	ggctctgtct	tttggtctcat	agcctttcat	420
catggcgatc	tattaaatga	tccagacatc	acaccttggt	ttgacaacgt	ccattctttc	480
acaggggcct	ttttgtttct	cctagagacc	caaaccacca	taggatattg	ttatcgctgt	540
gttactgaag	aatgtttctgt	ggccgtgctc	atgggtgatcc	tccagtccat	cttaagttgc	600
atcataaata	cctttatcat	tggagctgcc	ttggccaaaa	tggcaactgc	tcgaaagaga	660

```

gcccaaacca ttcgtttcag ctacttttga cttataggta tgagagatgg gaagctttgc 720
ctcatgtggc gcattggtga ttttcggcca aaccacgtgg tagaaggaac agttagagcc 780
caactttctc gctatacaga agacagtga gggaggatga cgatggcatt taaagacctc 840
aaattagtca acgaccaaata catcctggtc accccggtaa ctattgtcca tgaaattgac 900
catgagagcc ctctgtatgc ccttgaccgc aaagcagtag ccaaagataa ctttgagatt 960
ttggtgacat ttatctatac tgggtgattcc actggaacat ctcaccaatc tagaagctcc 1020
tatgtttccc gagaaattct ctggggccat aggttttaatg atgtcttggga agttaagagg 1080
aagtattaca aagtgaactg cttacagttt gaaggaagtg tggaggtata tgcccccttt 1140
tgcagtggca agcaattgga ctggaaagac cagcagctcc acatagaaaa agcaccacca 1200
gttcgagaat cctgcacgtc ggacaccaag gcgagacgaa ggtcatttag tgcagttgcc 1260
attgtcagca gctgtgaaaa ccctgaggag accaccactt ccgccacaca tgaatatagg 1320
gaaacacctt atcagaaaagc tctcctgact ttaaacagaa tctctgtaga atcccaaattg 1380
tagtctctaa ttgcaattat gagggctacc actgaatcat tttatctttc agccaatcaa 1440
gtcgttgtaa acgtggcttt tttgaaagtg ttatggctat gttttatgat gatgctgggt 1500
aagtagagta agttaaacctt ggtaaaagat aatctaaaaa ttccatagtt ctcagttatt 1560
aaaatttttc ttgttcgcca attttgtatt aagaatgcta ttaagcctaa ttgattaaaa 1620
tttatctttt ttattatctt acatgcttgt atcttcagtt ggaggtgtag tattcaaaaa 1680
cggggaatga aggcaggaag gaggctggaa taaataaaaa taaaatgatt 1730

```

&lt;210&gt; 43

&lt;211&gt; 1147

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1670595CB1

&lt;400&gt; 43

```

gcagctgtct tttccggccc ccgtgcactc tccgcccag gcgagcccc cggtctcgcg 60
ggatcgcccc cgagcgctgc gtcctgcggg tgggtcacct aaccatttg tggcttccct 120
tacctgtgct cagccatggc cagcgagagc tcacctctgc tggcctaccg gtcctgggg 180
gaggaggggg ttgccctccc tgccaatggg gccgggggtc ctggaggggc gctcgccgg 240
aagctgtcca ccttcttggg tgtgggtggg ccactgtcc tgtccatgtt cagcatagtt 300
gtttttctga ggattgggtt cgtgggtggg catgctgggc tactgcaggc cctggccatg 360
ctgctggttg cctacttcat cctggcactc accgtcctct ctgtctgtgc catcgccacc 420
aatggagccg tgcagggggg cggagccctac tgtatcctcc aacatcgatg gactgggatg 480
ccacagggcc cagtgggtcc cgggtcctgc ccagggtc cggcttggaa cctgctgtat 540
ggctccctgc tgctgggccc tgtgggtggg gtctgcacct tgggagccgg cctctatgcc 600
cgggcctcat tctcacatt cctgctgggt tctggctccc tggcctctgt gctcatcagt 660
tttgtggctg tggggccgag ggacatccgc ttgactccta ggccctggcc caatggctcc 720
tccctgcgcg cccggtttgg ccacttcacc ggcttcaaca gcagtacct gaaggacaac 780
ttgggcgctg gctatgctga ggactacacc acgggagccg tgatgaattt tgccagctc 840
tttgcgtgcc tctttaacgg caggcatcat ggctggggcc aacatgtcag gggagctgaa 900
ggaccccagc cgggcgatcc ctctgggcac gatcgctgcc gtcgectaca ccttcttctg 960
ctatgccctg cttttctttc tctccagcct cccttcactg gtgccttgat gctagggggc 1020
aggcctctc tgtgactctg ggctacctca gtttcccat tttggccaga ctcaccggcc 1080
caccggggtg gtgatgtttt cgttctgttt tatttttcta actctgcatg accatgaata 1140
aaagacc 1147

```

&lt;210&gt; 44

&lt;211&gt; 2745

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1859560CB1

&lt;400&gt; 44

```

cggcgacgcc agggacccca cgcataccga gtgaagcaac tagaactcca gggctgtgaa 60
agccacaggt gggggctgag cgaggcgtgg cctcaggagc ggaggacccc ccactctccc 120
tcgagcgccg cagtcaccgc tagcgggtgg agccgcctt ggtgcgcagt tggaaaacct 180
cggagccccc ctggatctcc tggctgccac ccgaccccc cgcagccta cgcaccaccg 240
tagagatgcc ttcttcgggtg acggcgctgg gtcaggccag gtccctctgg cccgggatgg 300
ccccgagcgc ctgctgctgc tcccctgcgg ccctgcagag gaggtgccc atcctggcgt 360
ggctgcccag ctactccctg cagtggctga agatggattt cgtcgccggc ctctcagttg 420

```

```

gcctcactgc cattccccag gcgctggcct atgctgaagt ggctggactc cgcgccagct 480
atggcctcta ctctgccttc atgggctgct tctgttatctt ctctctgggc acctcccggtg 540
atgtgactct gggccccacc gccattatgt cctctctggt ctctctctac accttccatg 600
agcccgcccta cgctgtgctg ctggccttcc tgtccggctg catccagctg gccatggggg 660
tcctgcgttt ggggttctctg ctggacttca ttctctaccc cgtcattaaa ggcttccacct 720
ctgctgctgc cgtcaccatc ggctttggac agatcaagaa cctgctggga ctacagaaca 780
tccccaggcc gttcttctctg cagggtgtacc acaccttctt caggattgca gagaccaggg 840
taggtgacgc cgtcctgggg ctggtctgca tgcctctggg gctggtgctg aagctgatgc 900
gggaccacgt gcctcccgtc caccocgaga tgccctctgg tgtgcggctc agccgtgggc 960
tggctctgggc tgccacgaca gctcgcaacg ccttggtggt ctctctcgca gccctgtgtg 1020
cgtactcctt cgaggtgact ggataccagc ctttcatcct aacaggggag acagctgagg 1080
ggctccctcc agtcgggac cgcctcttct cagtgtaccac agccaacggg acgatctcct 1140
tcaccgagat ggtgcaggac atgggagccg ggctggccgt ggtgccccctg atgggcctcc 1200
tggagagcat tgccgtggcc aaagccttcg catctcagaa taattaccgc atcgatgcca 1260
accaggagct gctggccatc ggtctcacca acatgttggt ctccctcgtc tcctctacc 1320
cggtcacagg cagctttgga cggacagccg tgaacgctca gtcgggggtg tgcaccccg 1380
cggggggcct ggtgacggga gtgctggtgc tgctgtctct ggactacctg acctcactgt 1440
tctactacat ccccaagtct gccctggctg ccgtcatcat catggccgtg gccccgctgt 1500
tcgacaccaa gatcttcagg acgctctggc gtgttaagag gctggacctg ctgcccctgt 1560
gogtgacctt cctgctgtgc ttctgggagg tgcagtacgg catcctggcc gggggccctg 1620
tgtctctgct catgctcctg cactctgcag ccaggcctga gaccaagggtg tcagaggggc 1680
cgggtctggt cctgcagccg gccagcggcc tgctcctccc tgccatggag gctctgagg 1740
aggagatcct aagccgggcc ctggaagtgt ccccggcagc ctgcccgtg ctggagtga 1800
cccatgtctg cagcatcgac tacactgtgg tgctgggact cggcgagctc ctccaggact 1860
tcagaagca gggcgtcgcc ctggcctttg tgggcctgca ggtccccgtt ctccgtgtcc 1920
tgctgtccgc tgacctgaag gggttccagt acttctctac cctggaagaa gcagagaagc 1980
acctgaggca ggagccaggg acccagccct acaacatcag agaagactcc attctggacc 2040
aaaagggtgc cctgctcaag gcataatggg gccacccgtg ggcateccaca gtttgagg 2100
tgttccggaa ggttcttctg actgtgatg gatgctggat gccgctgat agacatgctg 2160
gcctggctga gaaacccctg agcaggtaac ccagggaaga gaaggaagcc aggcctggag 2220
gtccacggca gtgggagtgg ggctcactgg ctctctgtgg gatgactgga aaatgacctc 2280
gctgctgttc cctggcatga ccctctttgg aagagtgggt tggagagagc cttctagaat 2340
gacagactgt gcgaggagc aggggcaggg gtttccagcc cgggctgtgc gaggcatcct 2400
ggggctggca gcacctccc ggctcaccag tgccacctgc gggggaggga gggggcaggc 2460
aggagtctgg gaggcgggtc cgctcctctt gtctgcggca tctgtgctct ccgagagaaa 2520
accaagggtg gtcaaatgac gtcaagtctc tatttaaaaa taattttgtg ttttctaaat 2580
ggaaaaagtg atagctttgg tgattttgta aaagtcataa atgcttattg taaaaaatac 2640
aggaaaccac cctcacctt gtccacttgg agatgcattc cagaccctc cccaaacatg 2700
catatgtacc tgtccgtcag tgtgtggatg tatgtttaca gttct 2745

```

&lt;210&gt; 45

&lt;211&gt; 3204

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5530164CB1

&lt;400&gt; 45

```

cgacctctgg agctactgag cctgcaagcc cagcctctct gcgcccaggg ctgcccggggc 60
agctggcgcc gcacaaatac ggggcgggac acggggcggg gtcccggggg 120
agggcctgag ccgcacagcc cggccagggg tgggtgcgtg aaacgggcgt ctggatcccc 180
gaatgggttc gtgtttccgt gtgtgggtcc gggggaggcc cacgaacgcc agcgaacccg 240
ctgacaccac cgcccaacta tgaactcatc aggcgcctga agaccgacac gccgaacatg 300
cgccgcgcgc actcgcgcac gagttagatc atcgcgcctc ggtcgtgagt gcgctcacac 360
gcagcctgag actcgacggg agggggtcac gtggaagtat ctgagagagg cgtacttggc 420
cactaggaaa gcacctccc ctttccaaaa atgctccgga agtgccctcg cctcgcaaac aagttcggcg 540
agatggccgg ggcagtcggc acgaggagg cggggatgcg cctaaagctta aggacctggc atttctcaag 600
gggaagatgg cggatgacaa ggattctctg cctaagctta gaagacgaag tcaacagtgg agtgggccag 660
aaccagctgg aaagcctgca gcggcgtgta cccgttcttc aggggatcc tggctggcta tgggtggcg 720
gatggctcgc tttgtctc catcagcagt attgggcttt gctgtgggca cctgcactgg catctatgag 780
gctcaggcat atgctgtgcc caacgtggag aagacattaa gggactatct gcagttgcta 840
cgcaaggggc ccgactagct ctaggtgcca tgggaagggc aggatgagca gctcagcctt 900
caggtggaga cactttatct ggattcccca gctgtcatcc atttgctatc tccaacttcc 960
ctgccacctt catccttgcc tcccttctg cagattgtgg acagtatgct ctccagcctg 1020

```

accctggatt	ccttcttccc	cttctagct	ccatgggact	cgcccccaaga	ctgtggcttc	1080
aaggaccacc	agcccccttac	tcttcaagcc	ctgactgtgg	agttggtaga	tgccctctgat	1140
cctcagtatt	ctctctggca	atgttccacg	gcttctcctt	cctgggagct	ggctccataa	1200
cttgattttc	cccaaacgtg	ttgcaatccc	tgtgccccct	tagccaccca	gggtcttgtg	1260
tgggtatgag	tgtagaggat	gggggtatgc	caggcctggg	ccgtcccagg	caggcccgtc	1320
ggaccctgat	gctactccta	tccactgcca	tgtacgggtg	ccatgcccc	ttgctggcac	1380
tgtgccatgt	ggacggccga	gtgcccttcc	ggccctcctc	agccgtgctg	ctgactgagc	1440
tgaccaagct	actgttatgc	gccttctccc	ttctggtagg	ctggcaagca	tggccccagg	1500
ggccccacc	ctggcgccag	gctgctccct	tcgcactatc	agccctgctc	tatggcgcta	1560
acaacaacct	ggtgatctat	cttcagcggt	acatggaccc	cagcacctac	cagggtgctga	1620
gtaatctcaa	gattggaagc	acagctgtgc	tctactgcct	ctgcctccgg	caccgcctct	1680
ctgtgcgtca	gggggttagcg	ctgctgtctg	tgatggctgc	gggagcctgc	tatgcagcag	1740
ggggccttca	agttcccggg	aacaccttcc	ccagtcccc	tccagcagct	gctgccagcc	1800
ccatgcccc	gcatatcact	ccgctaggcc	tgctgtccct	cattctgtac	tgccctcatct	1860
caggcttgte	gtcagtgta	acagagctgc	tcatgaagcg	acagcggctg	ccccctggcac	1920
ttcagaacct	cttccctctac	acttttgggtg	tgcttctgaa	tctaggctctg	catgctggcg	1980
gcggtctctg	cccaggcctc	ctggaagggt	tctcaggatg	ggcagcactc	gtgggtgctga	2040
gccaggcact	aaatggactg	ctcatgtctg	ctgtcatgaa	gcatggcagc	agcatcacac	2100
gcctctttgt	gggtgtctgc	tcgctgggtg	tcaacgccgt	gctctcagca	gtcctgctac	2160
ggctgcagct	cacagccgcc	ttcttccctgg	ccacattgct	cattggcctg	gccatgcgcc	2220
tgtactatgg	cagccgctag	tccctgacaa	cttccaccct	gattccggac	cctgtagatt	2280
ggcgccacc	accagatccc	cctcccaggc	cttctccct	ctcccatcag	cagccctgta	2340
acaagtgcct	tgtgagaaaa	gctggagaag	tgagggcagc	cagggtattc	tctggagggt	2400
ggtggatgaa	gggggtacccc	taggagatgt	gaagtgtggg	tttgggttaag	gaaatgctta	2460
ccatccccc	cccccaacca	agttcttcca	gactaaagaa	ttaaggtaac	atcaatacct	2520
aggcctgaga	aataaccccc	tccttgttgg	gcagctccct	gctttgtcct	gcatgaacag	2580
agttgatgaa	agtgggggtg	gggcaacaag	tggttttct	tgccctactt	agtcacccag	2640
cagagccact	ggagctggct	agtcagccc	agccatgggt	catgactctt	ccataaggga	2700
tcctcaccct	tccactttca	tgcaagaagg	cccagttgcc	acagattata	caaccattac	2760
ccaaaccact	ctgacagtct	cctccagttc	cagcaatgcc	tagagacatg	ctccctgccc	2820
tctccacagt	gctgtctccc	acacctagcc	tttgttctgg	aaaccccaga	gagggctggg	2880
cttgactcat	ctcaggggaat	gtagccctgc	ggccctgggt	taagccgaca	ctcctgacct	2940
ctctgttcac	cctgagggct	gtcttgaagc	ccgtacccca	ctctgaggct	cctaggaggt	3000
accatgcttc	ccactctggg	gcctgcccct	gcctagcagt	ctcccagctc	ccaacagcct	3060
ggggaagctc	tgcacagagt	gacctgagac	caggtagagg	aaacctgtag	ctcaatcagt	3120
gtctctttta	ctgcataagc	aataagatct	taataaagtc	ttctaggctg	taggggtgggt	3180
cctacaacca	cagccaaaaa	aaaa				3204

&lt;210&gt; 46

&lt;211&gt; 2763

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 139115CB1

&lt;400&gt; 46

tgcatttgc	atgactttga	ccggtccact	gacaacgcaa	tatgtttatc	ggagaatatg	60
ggaagaaact	ggcaactaca	ctttttcctc	tgatagcaat	atctctgagt	gtgaaaaaaa	120
caaaagcagc	ccaatttttg	cattccaggga	ggaagttcag	aaaaaagtgt	caggttttaa	180
tctgcagatg	gacataagtg	gattaattcc	tggtctagt	tctacattca	tacttttgtc	240
tattagtgat	cactacggac	gaaaattccc	tatgattttg	tcttccgttg	gtgctcttgc	300
aaccagcgtt	tggctctgtt	tgccttgcct	ttttgccttt	ccattccagc	ttttgattgc	360
atctaccttc	attggtgcat	tttgtggcaa	ttataccaca	ttttggggag	cttgccttgc	420
ctatatagtt	gatcagtgtg	aagaacacaa	acaaaaaaca	attcgaatag	ctatcattga	480
ctttctactt	ggacttggtt	ctggactaac	aggactgtca	tctggctatt	ttattagaga	540
gctagggttt	gagtggctgt	ttctaattat	tgctgtgtct	cttgcctgta	atttgatcta	600
tatttttatt	tttctcggag	atccagtgaa	agagtgttca	tctcagaatg	ttactatgtc	660
atgtagtga	ggcttcaaaa	acctatttta	ccgaacttac	atgcttttta	agaatgcttc	720
tggtaaagaga	cgatttttgc	tctgttttgt	actttttaca	gtaatcactt	atttttttgt	780
ggtaattggc	attgccccaa	tttttatcct	ttatgaattg	gattcaccac	tctgctggaa	840
tgaagttttt	ataggttatg	gatcagcttt	gggtagtggc	tcttttttga	ctagtttctc	900
aggaatatgg	cttttttctt	attgtatgga	agatattcat	atggccttca	ttgggatttt	960
taccacgatg	acaggaatgg	ctatgaccgc	gtttgccagt	acaacactga	tgatgttttt	1020
agccaggggtg	ccgttccctt	tcactattgt	gccattctct	gttctacggg	ccatgttgte	1080
aaaagtgggt	cgttcgactg	aacaagggtac	cctgtttgct	tgtatttgct	tcttagaaac	1140

```

acttggagga gtcactgcag tttctacttt taatggaatt tactcagcca ctgttgcttg 1200
gtaccctggc ttcactttcc tgctgtctgc tgggtctgta ctacttccag ccatcagctc 1260
atgtgttgtc aagtgtacca gctggaatga gggaagctat gaacttctta tacaagaaga 1320
atccagtga gatgcttcag acaggtgact gtgatttaaa caaacaaaaa aaatctatga 1380
atgcacatat catataccat gacttctgaa gactataaat gaattccaca atcagtgcct 1440
cactgagaac caattttacc tatcttttct tctaaactga acagtcagag agacagctcc 1500
tggcttttagc ttcttgtggt accacgcact ttgagcactt tgtgcgtatc atgcaatata 1560
cttgcaatac acagaacaaa tttcaaatac gcctcacttt tagacttaga agagaaacat 1620
taaaacttaa ggggtgtaagg agggatcaag aaacttgata aggtcaaaag caataatctc 1680
tctgacatat tccaggctct tacactgaga ccaaagagaa atctttacct cagtttcttc 1740
atcagcagaa tgggtttctg gcctctctca gggataattt tgaaggcata atgaaaatta 1800
tgatgaatca ctcatgtgta ggaaaataat gatataagtt tcaaatatgt atgattttac 1860
ctatacttg taatgctttg ttttatagag cctgttaagc tgctattgat agtcggagct 1920
tatatactgt gacttctgaa gactatacat gaattccaca atcagtgcct tgttgatata 1980
aaatccttaa aaggagggca ctttaaagaa tatgtatttt tcacttttct taatatgttt 2040
catcggtgac aggcattgata atatttctat atgtaatggg taattgggaa aaaatagatg 2100
ataaataaaa ttgctctaaa gaagttaaaa aactgaatga acagctaata ctggtataaa 2160
gtaactaatg tttggagcca acatttgttc cttgtgtcag caaaaggata ttcacattcc 2220
atgatccctg gctgagaatt ctgcctctag tctttcttac ccagctgttg tctatccttg 2280
ttcaattata aatactgcta agggcatttt taaaatacga tcttgtagtc cttaaatttg 2340
aatccgtcag cacggctcact cataggaaaa tgatcaaaaca agcaagccag tcatgatttg 2400
actccttccc atctcatttc ttactgcctt acgctcatcc tgagggtccac cttggtctct 2460
aaaaacacca tgtgtttctc tgctctcaca tctttctaca cactgttcca tttgtcttc 2520
ctccacatt acattgaaac tttcaagcct cagtcgaaac attgcttctt ctggtatgca 2580
gccttcttga catcctctct cactcccag tccctacagg gcttccatag ctcttttgtg 2640
gcacttcgat cccagcattt tccatcgact tgtaattgtt tctgctacct gacaatcatc 2700
gccttgagta ctgggacaac ctttgattac tcattatata ctcaataaat atttgttgaa 2760
cta

```

&lt;210&gt; 47

&lt;211&gt; 1639

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1702940CB1

&lt;400&gt; 47

```

atcgactga ggcttgagtc tgacttctct cccacactg ctgtgccctt aaactgcaga 60
gatcggggcg ggggttgggg ggcaagcggc tcatatgggt tcaaaaaact cccaggtctc 120
aactctggtt ctgactgcct gagacatggg gacgtgacac agcagacctt gaatcctgag 180
gatgtgaggc aggttatatc tgggaggcgg gaggacgtgt ctggttatta cacagatgca 240
cagctggacg tgggatccac acagctcaga acagttggat cttgctcagt ctctgtcaga 300
ggaagatccc ttggacaaga ggacctgccc ttggtgtgag agtgagggta gaggaagctg 360
gaacgagggg taaggaaaaac cttccagctc ggacagtgc ggacagtgc tggagagctc 420
ccctcggtaa ccagccgct ggcacctga accagagag cagtatcttt attgaggatt 480
accttaagta tttccaggac caagtgagca gagagaatct gctacaactg ctgactgatg 540
atgaagcctg gaatggattc gtggctgctg ctgaactgcc cagggatgag gcagatgagc 600
tccgtaaagc tctgaacaag cttgcaagtc acatggtcat gaaggacaaa aaccgccacg 660
ataaagacca gcagcacagg cagtggtttt tgaaagagtt tctcgggttg aaaaggggagc 720
ttgaggatca cataaggaag ctccgtgccc ttgcagga ggttgagcag gttccacagag 780
gcaccaccat tgccaatgtg gtgtccaaact ctgttggcac tacctctggc atcctgacct 840
tctcggcctt gggctctggca cccttcacag aaggaaatcag ttttgtgctc ttggacactg 900
gcatgggtct gggagcagca gctgctgtgg ctgggtgga gcaacttggg ccaaagcggc 960
taaacaattt gggggcacga gcccagccc gacacaccaa tgttcttacc ttagttgaca 1020
caaaggatgat gaaggagtgt gtgggtggga ggaacatccg tgccatcaga cgagccagag 1080
attggtacca agtcacacaa ggtattggga ggaacatccg cattggggcg atctcagctg 1200
ccaaccctca gttaggagcg tatgcccac ccccgcatgt ccaggcaatg agcagaggaa 1260
aaggcgggtg acaggttgag aggttgttg actggaggca gctggatgtg gtcagccttg 1320
ccatgatcgt ggggtgcagc cttgaggggg ggaagctca actttctcac gtcagctgag gagctgaaga 1380
catatgagtc aaagcacttg gggagctggag ggaagctca actttctcac caagatccat gagatgctgc 1440
agcgggctca agaccaatga cccagagca gtgcagccac cagggcagaa atgccgggca 1500
agccaggcca caaaatgcag actttttttt ttttcaagtc tttgacgggg aaggggagctc 1560
cagggcagga cccagtaggg gtggcggggc ccaactctgg gccgtgtgaa cctcccgggg 1620
ggggggatc gattaacgc

```

<210> 48  
 <211> 1600  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1703342CB1

<400> 48  
 caagggcgcc caggacaggc aggggctgca cgcgggtgaag aaaccaagac gcagagaggc 60  
 caagcccctt gccttgggtc acacagccaa aggaggcaga gccagaactc acaaccagat 120  
 ccagaggcaa caggggacatg gccacctggg acgaaaaggc agtcaccgc agggccaagg 180  
 tggctcccgc tgagaggatg agcaagttct taaggcactt cacggtcgtg ggagacgact 240  
 accatgectg gaacatcaac tacaagaaat gggagaatga agaggaggag gaggaggagg 300  
 agcagccacc acccacacca gtctcaggcg aggaaggcag agctgcagcc cctgacgttg 360  
 cccctgcccc tggccccgca cccaggggcc ccttgactt caggggcatg ttgaggaaac 420  
 tggtcagctc ccacagggtt cagggtcatca tcatctgctt ggtggttctg gatgccctcc 480  
 ttgtgcttgc tgagctcatc ctggacctga agatcatcca gcccgacaag aataactatg 540  
 ctgccatggt attccactac atgagcatca ccatcttggg cttttttatg atggagatca 600  
 tcttttaaatt atttgtcttc cgcctggagt tctttcacca caagtttgag atcctggatg 660  
 ccgtcgtggt ggtggtctca ttcatcctcg acattgtcct cctgttccag gaggaccagt 720  
 ttgaggctct gggcctgctg attctgctcc ggctgtggcg ggtggcccg atcatcaatg 780  
 ggattatcat ctcagttaag acacgttcag aacggcaact cttaaggtta aaacagatga 840  
 atgtacaatt ggccgccaa agttcaacacc ttgagttcag ctgctctgag aaggaaacag 900  
 aaattgaaag acttaacaaa ctattgcgac agcatggact tcttggtgaa gtgaactaga 960  
 cccggaccag cccccctcaa aaagaagaca ctgtctcatg ggctgtgct gtcacgagag 1020  
 gaacagctgc cctcctggg ccgcttggtg agaggtttgg tttgatacct ctgctccct 1080  
 cctgccagca tggattctgg gtggacacag ccttgtggaa ggtccagtac caccaagagc 1140  
 tgcccatcca cccccacccc acactgtatc aaatgtatca cattttctca tgttgaacac 1200  
 tttagcctta attgaaaatg agcaacaaag ctggacaatt gctagtgtga tataaaattt 1260  
 aatctcaccg aatgtacagt tttcaaattt cactgtgata ttaaggaact gatgcatctg 1320  
 agcattctga aagaaagaaa aagaagctac tttagctgcc accccattct agaaaaagtct 1380  
 cttatttttca agctgttcta aatagcttcg tctcagtttc cccaaaagggt gtaccaggc 1440  
 cctcctctg tgtgccccag ctgcatcagc cagcttctag gtggctccat tgttttctgc 1500  
 cacctgacaa ctttttctct caattactgt acaactactg tataaaataa aacaactact 1560  
 gtataaaata aactctctct tttccctgga aaaaaaaaaa 1600

<210> 49  
 <211> 2380  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1727529CB1

<400> 49  
 ctgagccatg gggggaaagc agcgggacga ggatgacgag gcctacggga agccagtcaa 60  
 atacgacccc tcttttcgag gccccatcaa gaacagaagc tgcacagatg tcatctgctg 120  
 cgtcctcttc ctgctcttca ttctagggtta catcgtgggt gggattgtgg cctgggtgta 180  
 tggagacccc cggcaagtc tctaccccag gaactctact ggggctact gtggcatggg 240  
 ggagaacaaa gataagccgt atctcctgta cttcaacatc ttcagctgca tctgttccag 300  
 caacatcctc tcagttgctg agaacggcct acagtgcctc acaccccagg tgtgtgtgtc 360  
 ctctgccccg gaggacccat ggactgtggg aaaaaacgag ttctcacaga ctgttgggga 420  
 agtcttctat acaaaaaaca ggaacttttg tctgccaggg gtacctgga atatgacggt 480  
 gatcacaagc ctgcaacagg aactctgccc cagtttctc ctcccctctg ctccagctct 540  
 gggacgctgc tttccatgga ccaacattac tccaccggcg ctcccaggga tcaccaatga 600  
 caccaccata cagcagggga tcagcgggtc tattgacagc ctcaatgccc gagacatcag 660  
 tgttaagatc tttgaagatt ttgcccagtc ctggtattgg attcttgttg ccttgggggt 720  
 ggtcttggtc ttgagcctac tgtttatctt gcttctgcgc ctgggtggctg ggcccctggt 780  
 gctggtgctg atcctgggag tgetgggcgt gctggcatac ggcactctact actgctggga 840  
 ggagtaccga gtgctgcggg acaagggcgc ctccatctcc cagctgggtt tcaccaccaa 900  
 cctcagtgcc taccagagcg tgcaggagac ctggctggcc gcctgatcg tgttggcggt 960  
 gcttgaagcc atcctgctgc tgggtgctcat ctctctggcg cagcggattc gtattgccat 1020  
 cgccctcctg aaggaggcca gcaaggctgt gggacagatg atgtctacca tgttctacce 1080  
 actgggtcacc tttgtctctc tctcctctg cattgcctac tgggcatga ctgctctgta 1140

cctggctaca	tcggggcaac	cccagtatgt	gctctgggca	tccaacatca	gtccccccgg	1200
ctgtgagaaa	gtgccaataa	atacatcatg	caacccccacg	gccacacctg	tgaactcctc	1260
gtgcccagg	ctgatgtg	tcttccagg	ctactcatcc	aaaggcctaa	tccaacgttc	1320
tgtcttcaat	ctgcaaatct	atggggtect	ggggtctctc	tggaccctta	actgggtact	1380
ggccctgggc	caatgcgtcc	tcgctggagc	ctttgcctcc	ttctactggg	ccttccacaa	1440
gccccaggac	atccctacct	tccccttaat	ctctgccttc	atccgcacac	tccgttacca	1500
cactgggtca	ttggcatttg	gagccctcat	cctgaccctt	gtgcagatag	cccgggtcat	1560
cttggagtat	attgaccaca	agctcagagg	agtgcagaac	cctgtagccc	gctgcatcat	1620
gtgctgtttc	aagtgcgtcc	tctgggtgtc	ggaaaaattt	atcaagttcc	taaaccgcaa	1680
tgcatacatc	atgatcgcca	tctacgggaa	gaattttctgt	gtctcagcca	aaaatgcgtt	1740
catgctactc	atgcgaaaca	ttgtcagggt	ggctgtcctg	gacaaagtca	cagacctgct	1800
gctgttcttt	gggaagctgc	tgggtggtcg	aggcgtgggg	gtcctgtcct	tctttttttt	1860
ctccggctgc	atcccggggc	tgggtaaaga	ctttaagagc	ccccacctca	actattactg	1920
gctgcccata	atgacctcca	tccctggggc	ctatgtcatc	gccagcggtc	tcttcagcgt	1980
tttcggcatg	tgtgtggaca	cgctcttctc	ctgcttctct	gaagacctgg	agcgggaaca	2040
cggtcccttg	gaccggccct	actacatgtc	caagagcctt	ctaaagattc	tgggcaagaa	2100
gaacgaggcg	cccccgga	acaagaagag	gaagaagtga	cagctccggc	cctgatccag	2160
gactgcacc	cacccccacc	gtccagccat	ccaacctcac	ttcgctttac	aggtctccat	2220
tttgtggtaa	aaaaagggtt	taggccaggc	gccgtggctc	acgcctgtaa	tccaacactt	2280
tgagaggctg	aggcgggctg	atcacctgag	tcaggagttc	gagaccagcc	tggccaacat	2340
ggtgaaacct	cgtctctat	taaaaatata	aaaattagcc			2380

&lt;210&gt; 50

&lt;211&gt; 3038

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2289333CB1

&lt;400&gt; 50

aggggcagg	aggcgggcac	caggcgcggg	tcctccggg	caggcgagg	aggcctgggc	60
ctgacgccg	ccacgcagcg	gcgggagagt	gagcactcgg	gcggcgccgt	cctggagacc	120
cgcgagagat	ggaagcgcg	gcgacgccgg	cggtgcggg	ggcgggcagg	cgcgaggagc	180
tagatatgga	tgtaatgagg	cccttgataa	atgagcagaa	ttttgatggg	acatcagatg	240
agaacatga	gcaagagctt	ctgcctgttc	agaagcatta	ccaacttgat	gatcaagagg	300
gcatttcatt	gtacaaaact	cttatgcacc	ttcttaaagg	aaatattgga	actggccttt	360
taggacttcc	attggcaata	aaaaatgcag	gcatagtgct	tggaccaatc	agccttgtgt	420
ttataggaat	tatttctgtt	cactgtatgc	acataattgg	acgttgcagt	cactttctat	480
gtctgagggt	taaaaagtca	acattagggt	atagtgacac	tgtgagcttt	gctatggaag	540
tgagtctctg	gagttgtctt	cagaagcaag	cagcatgggg	gcggagtgtg	gttgactttt	600
ttctgggtgat	aacacagctg	ggattctgtg	gtgtttatat	tgtcttctta	gctgaaaatg	660
tgaacaagt	tcatagaagga	ttcctggaga	gtaaagtgtt	tatttcaaat	agtaccaatt	720
catcaaacc	ttgtgagaga	agaagtgttg	acctaaggat	atatatgctt	tgttttcttc	780
catttataat	tcttttggtc	ttcattcgtg	aactaaagaa	tctatttgta	ctttcattcc	840
ttgccaacgt	ttccatggct	gtcagtcctg	tgataattta	ccagtatgtt	gtcaggaaca	900
tgccagatcc	ccacaacctt	ccaatagtgg	ctgggtggaa	gaaataccca	ctcttttttg	960
gtactgctgt	atttgctttt	gaaggcatag	gagtggctct	tccactggaa	aaccaaata	1020
aagaatcaaa	gcgtttccct	caagcggtga	atattggcat	ggggattgtt	acaactttgt	1080
atgtaacatt	agctacttta	ggatatatgt	gtttccatga	tgaatcaaa	ggcagcataa	1140
ctttaaatct	tccccaagat	gtatgggttat	atcaatcagt	gaaaattcta	tattcctttg	1200
gcatttttgt	gacatatcca	attcagttct	atgttccagc	agagatcatt	atccctggga	1260
tcacatccaa	atttcatact	aaatggaagc	aaatctgtga	atttgggata	agatccttct	1320
tggttagtag	tacttgtgcc	ggagcaatcc	ttattcctcg	tttagacatt	gtgatttctc	1380
tcgttggagc	tgtgagcagc	agcacattgg	ccctaactct	gccacctttg	gttgaaattc	1440
ttacattttc	gaaggaacat	tataatatat	ggatgggtct	gaaaaatatt	tctatagcat	1500
tcactggagt	tgttggcttc	ttattaggta	catatataac	tgttgaagaa	attatttatc	1560
ctactcccaa	agttgtagct	ggcactccac	agagtccctt	tctaaatttg	aattcaacat	1620
gcttaacatc	tggtttgaaa	tagtaaaagc	agaatcatga	gtcttctatt	tttgtcccat	1680
ttctgaaaa	tatcaagata	actagtaaaa	tacatttgcta	tatacataaa	aatggtaaca	1740
aactctgttt	tctttggcac	gatattaata	ttttggaagt	aatcataact	ctttaccagt	1800
agtggtaa	ctatgaaaaa	tccttgcctt	taagtgttag	caatagtcca	aaaaataag	1860
ttctgaaaa	tgaaaaaatt	aaaatgtaaa	aaaattaaag	aataaaaaata	cttctattat	1920
tcttttatct	cagtaagaaa	taccttaacc	aagatatctc	tcttttatgc	tactcttttg	1980
ccactcactt	gagaacagaa	taggatttca	acaataagag	aataaaataa	gaacatgtat	2040
aacaaaaagc	tctctccaga	tcattccctgt	gaatgccaaa	gtaaacttta	tgtacagtgt	2100



aaaaaaaaa	aatctcagtt	atgtttttat	tagccaaatt	ctaattgattg	gctcctggaa	2160
gtatagaaaa	ctcccattaa	cataatataa	gcatacagaaa	attgcaaaca	ctagaattaa	2220
ttttacactc	taatggtagt	tgatcttcat	agtcaagagg	cactgttcaa	gatcatgact	2280
tagtgtttca	atgaaatttg	acaaggggac	ttaaaactta	tccagtgcga	ctcccttggt	2340
tttcgtcaga	ggaaaaggag	gcctagaaag	gttaagtaac	ttgggtcgaga	ccactcagcc	2400
ttgagatcaa	gaaaacctaa	tcttctgact	cccaggccag	gatgttttat	ttctcacatc	2460
atgtccaaga	aaaagaataa	attatgttca	gcttaatttt	agtgttgaat	ctatttgatt	2520
atattttaat	actttgaaaa	tgaatgtgtg	atttttaata	gtatatgtga	cctgagcaga	2580
aaatcaggga	actccaagaa	gcctacactg	tggccatata	aacctcagca	agagaaagaa	2640
gctatgttct	tttaaaacag	aatagagacc	gcttgctggg	gaaactcctg	gctagtaaga	2700
tgtgtgtcta	gctatactat	ttgtggcttg	agctttttta	attattacct	tcctttcctg	2760
agttttgtag	gcaccacatt	cctgaatggc	agaaaataga	cacctcagaa	aacggaggat	2820
ttgtggactc	tttccagccc	tgtggctttt	cttatcacag	ccttttattt	attatgagca	2880
gaataaaaga	atcagctagg	tgtggtgggt	tgtgcttata	atcccagcta	ctctggagga	2940
taagttggga	ggatcacttg	agaggccagg	agcttgagac	cagcctgggc	agcatagtga	3000
gacctcgact	ctataaaaca	taaaaaaaaaa	aaaaaaaaa			3038

&lt;210&gt; 51

&lt;211&gt; 2608

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2720354CB1

&lt;400&gt; 51

taggctaatt	ttttttacag	acacgatttc	gccacgttgg	ccaggctggg	cttgaactcc	60
tgacctcaag	tgatccaccc	acctcagcct	cccaaagtgt	tgggattaca	ggcgtgagcc	120
actgcacctg	gccaggtcca	tcactttttg	cgcctattgc	ctogaagcca	gtctctgatg	180
ggacattagg	gcagggggccc	ttcagcctag	tctggggacat	gggcccgtca	ctcagcagta	240
tgacaagcat	cacctggaga	acggggccagt	ctcaggagggt	cgttcatgcc	ccactggcag	300
tgcaactgtg	agccatagtg	taaacaagag	gcttaacctg	aactgggtctg	agatcttggg	360
gacccctac	cctgtctcca	gcagcctgtc	ccttttagctg	tttgccctact	ggcaccctcat	420
cctgagaagg	catagatacc	cggcccaccc	tgccttgga	ttacaaaagt	cttagactgt	480
gcctgagtgc	ccggcctcct	tgggagaccc	tcctaggcag	cctaagcacc	agaccgggga	540
gctgggtgct	ctggtctctg	ctgcctgcct	ctcactggac	cctctccttc	caggtacggc	600
ttcaggtcca	gagcgtggag	aagcctcagt	accgcgggac	gttgcaactgc	ttcaagtcca	660
tcataagca	agagagcgtg	ctgggcctgt	acaagggcct	gggctcgccg	ctcatggggc	720
tcaccttcat	caacgcgctg	gtgttcgggg	tgcagggcaa	cacctccggg	gccctggggc	780
acgactcgcc	cctcaaccag	ttcctggcag	gtgcggcgcc	gggcccacatc	cagtgcgtca	840
tctgtgcgcc	catggagctg	gccaagacgc	ggctgcagct	gcaggacgcg	ggccgcggcg	900
gcacctacaa	gggctcgctg	gactgcctcg	cgagactcta	cgggcacgag	ggctctgcgtg	960
gcgtcaaccg	gggcatgggtg	tcacagttgc	tgcgtgagac	tcccagcttc	ggcgtctact	1020
tcctcaccta	tgacgctctc	acgcggggcg	tgggctgcga	gccggggcgac	cgctgctggg	1080
tgcccaagct	gctgttggcg	ggcggtacgt	caggcatcgt	gtcctggctc	tctacctate	1140
ctgtggagct	ggtagtcg	cggctgcagg	cggacggact	gcgggggcgcc	cgcgcttacc	1200
gcggcatcct	ggactgcgtg	caccagagct	accgcggcga	gggctggcgc	gtcttccacac	1260
gggggctggc	gtccacgctg	ctgcgcgcct	tcccctgcaa	cgtgccacc	ttcgccaccg	1320
tcacgggtgt	gtcacctac	gcgcgcggcg	aggaggccgg	gcccagaggc	gaggctgtgc	1380
ccgcgcgcgc	tgcggggcct	gccctggcgc	agccctccag	cctgtgacgc	tcaccccgcc	1440
ctccttcccc	agggtcctt	ctcagaaaacc	tgggacataa	attggccctt	gagtcgattg	1500
ccctgcttcc	tgttgggatg	ctgcgagctg	tggagtctat	cagatgtggg	ctgaattttg	1560
ctgatcagct	gggtagtttt	ggccgagaac	tgcacttgcc	tcagtgttct	catctatgaa	1620
ataaggaccc	tcatgcccac	actgtagagt	cacgaagctc	agagattatt	cccagcagca	1680
gccagcacct	ggcctggctg	aggccattgc	accgttatcc	tggaaactga	ggcagacact	1740
ccagccctct	tctgggatcc	tggccacgtc	attgtgctcc	tgcctgcag	gctggctccc	1800
gggggtctct	gatggccaac	caagggggcca	cccaggggacc	tctaactcca	cacatcctcc	1860
acccgggggg	gtgtgtgggc	acccctcttg	tctgtgttag	ggacagagga	aaacttgggtg	1920
tgcctccttg	tgtcacagaa	ctggatcctc	tgcatacccc	agcttctcca	catgccactg	1980
ctaggggtac	cccagctgct	gccactcctg	ctggagggtg	aactggggac	cctgcaccct	2040
ccgggaagcc	atggagtctg	ctggaggcac	catatcagcc	tgcgggacta	gggtggggag	2100
caaacaggcc	agcgggtggag	gtctggacag	ttcaagtgtg	atgcagctgt	ggcaaggaga	2160
aatccttccg	cctctgggccc	tcaggctgcc	tgtccataaa	atggggacat	ggccagctga	2220
cggacaactg	agtctccggc	ccacctacca	ccgccagcca	ggatccccc	aagtgtgcag	2280
agggctcagc	agagaacagt	atgggacccc	ctcaccaggc	ctggaacacc	tccagccaca	2340
aagaagccaa	aggtcagtec	ctctgctccc	cagcaaacgg	tgcctcccag	gcattctcag	2400

tgccagggt	tcattccctgt	gaaggcacag	ggcctgctag	tgggcacagg	ggtggctagt	2460
tggggcctgg	ggcagaggag	ggctgcacca	ggcgtcctgg	ggaatgtgct	cagtgaagac	2520
gacactgggc	tttgcacagc	ctgggtgtgc	tgtacagaaa	ctgtcaaggg	aataaagtgt	2580
tctttgtttt	ttaaaaaaaa	aaaaaaaa				2608

&lt;210&gt; 52

&lt;211&gt; 3804

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3038193CB1

&lt;400&gt; 52

ccctttcctg	tcactggcta	ctaccactcc	caaccctcct	caaagccgcc	ggagcaaccc	60
ccagggtcttt	actttacaat	cggcaatttg	acttgcctctg	ctgcatgtct	ggaggggacca	120
aggaaagtgt	ggagacgctc	caaggattag	gtgatcggag	cttgaaaaga	aaaaaagcca	180
aacaaataaa	caaaacccac	ccaccctaac	aaatatgagg	ctgctggaga	gaatgaggaa	240
agactgggttc	atggtcgga	tagtgctggc	gatcgctgga	gctaaactgg	agccgtccat	300
aggggtgaat	gggggaccac	tgaagccaga	aataactgta	tcctacattg	ctgttgcaac	360
aataattcttt	aacagtggac	tatcattgaa	aacagaggag	ctgaccagtg	ctttgggtgca	420
tctaaaactg	catcttttta	ttcagatctt	tactcttgca	ttcttccag	caacaatatg	480
gctttttctt	cagcttttat	caatcacacc	catcaacgaa	tggcttttaa	aagggttgca	540
gacagtaggt	tgcattgcctc	cgctgtgtc	ttctgcagtg	attttaacca	aggcagttgg	600
tggaaatgag	ggcatcggtta	taacacccct	gctcctgctg	ctttttcttg	gttcattctc	660
ttctgtgcct	ttcacatcta	ttttttctca	gctttttatg	actgttgtgg	ttcctctcat	720
cattggacag	attgtccgaa	gatacatcaa	ggattggcct	gagagaaaga	agcctccttt	780
tgggtgctatc	agcagcagtg	tactcctcat	gatcatctac	acaacattct	gtgacacgtt	840
ctctaaccaca	aataattgacc	tggataaatt	cagccttggt	ctcatactgt	tcataaatatt	900
ttctatccag	ctgagtttta	tgccttttaac	tttcatcttt	tcaacaagga	ataattcggg	960
tttcacacca	gcagacacag	tggctatcat	tttctgttct	acacacaaat	cccttacatt	1020
gggaattccg	atgctgaaga	tctgttttgc	aggctatgag	catctctctt	taatatctgt	1080
acccttgctc	atctaccacc	cagctcagat	ccttctggga	agtgtgttgg	tgccacaacat	1140
caagtcttgg	atggtatcaa	ggcagaagaa	actactccaa	accagggggc	cactggctaa	1200
cttgaataat	ccagaaggct	tggaaatatc	atccatcaaa	tttgggcatt	aaaataaata	1260
ccaagagtcc	atcctccagg	gagtgaagct	gacaaaggcc	acagtataac	aaaggagggtg	1320
gactttctgt	agcaatgtat	atatgtacag	gattgtacat	actagcaatt	ctgaagactt	1380
gtacttgtga	atgttgcttc	aatgcatatt	ttattttttt	acacaaaaat	atgagatcct	1440
gtttaagtgc	cttaaaatgt	atttgacaag	agcgttattt	ccacaatatg	ctttgttgat	1500
tactgccagg	gggtgttaca	tatttggggg	tttaattttg	tttctaatg	caggaatcag	1560
tcattgtaag	tgacaaaaag	caaacatgct	ttcctgcag	cacctttgtg	taatacaacc	1620
ctatagttagt	tactgtaatg	tttgaaatga	ggtcacacca	tcaggaaaat	gcccttctga	1680
tgacagtga	aattttccaaa	gtcttattca	tgcatacttt	gatttactgt	gtgattcttt	1740
ttttctacga	ctgtgacatg	cctcttcctt	atcaactcag	caggggtcat	agatcgaata	1800
gatgtgaaa	agcgttaagt	atatgcattc	cttgacatca	tttttaaaga	cattccttca	1860
aatagtttcc	acacagaaat	tcctcactcc	cattgtgaga	gattgtgggt	atatgtctta	1920
aatttattat	aagctgtctc	aaagaaaggg	tctgaatgtt	tgaattatga	gtgaaatcat	1980
gtgaaatttt	gagttaaact	ctgtgatttg	attttcaggg	tctttaaaat	atatcttaat	2040
atcttcttcc	tctttattca	ataatttctg	tcttgcaact	acacactcat	aacagccaaa	2100
tatgaggcac	aaaaatgtta	caatcagttt	gaaagcagca	tcaattaatg	gtagattcta	2160
ttcacattcc	acaacccaga	ccaaattttt	ttcctattac	gcagatgtgc	tgagcacttt	2220
ccagattgcc	cctgttggcc	aaaagcagcc	tgttacatcc	tggaaatga	cacacttaag	2280
gtatttgaga	caattttatta	atgaaaattt	ccttggcaga	tttgacaaat	gttggcaata	2340
tttttttaaa	agttaaatca	tattgtcttc	atgaataaat	gaaaatataa	agggtcatgga	2400
tgcaaaacaaa	tgttacatat	acacattctg	tctctccaga	tgaaaagaac	atgcaaaacc	2460
atttaataaac	caaaaatatca	agtaaaaatta	gttcccaacg	gggcagcagc	tttcaaatga	2520
gtgtccaata	tttgcttctg	ctatagctgc	aagaactgta	actggaccca	agtagagaat	2580
gaagccacgt	atagaactac	gagaacactt	ttctgtgttt	cccccatgcc	gtcctgtcac	2640
atcctctttac	acgtcctctc	ttgatttgat	agacaatatt	ggcatcctgg	gtctcactga	2700
ggcgtgtcta	tgtcctcagc	agctgttttt	gttgtttcgt	tattatgccc	acaacaaaaa	2760
atcattctctt	agaaattcac	caagtttatc	tactgtgtaa	atttatatta	ttgttactac	2820
cagggtctcat	cttttgtcaa	tgtcattgaa	taaatttcat	aagagttatt	ctcagtgatga	2880
attttaaggc	taatgccaga	tcttgcaaaa	atctatgcta	accaggctgt	agtacacact	2940
gtttataaaga	attttacttg	tgtctaaaac	tacagtaatt	ttgcttaggt	aattgtgctt	3000
acctatggag	cacaggaagg	ctcttaggtt	ttgttcctac	aagtttcttt	gaatttttga	3060
gtaaatggaa	gtgtctgtct	gtctgtcctc	tatctgcctt	atcataaaaa	tctttctccc	3120

```
taacattaata atactgatcc ccgcccccaa cttatctacc tctattgtct aacacctata 3180
gtagggtgtga tcatgggata aaattcaact gaaaatgcta tgataacatt ttatcgtttg 3240
ctttaaaaaat gtgctttgtt ttcaaataat ctttacatag tgaactttgg tggcgtagt 3300
gatatgttta tgcctatttc ttttttttac acaaattcct tggcatattt ttccataaag 3360
aacaaaaaat aaaatcaaaa tttattttta attcatgctt attgggattt aattatttcag 3420
agcttaaaat attttgttat gtttatacac tgtaaagcta tctgttttat gcatttgttt 3480
tgtctaaatg tatttatgaa agaaatacat tagattatat ttatgtttac tcattttttcc 3540
acctggattt tttttaatgg ttgttacaaa attagatttt ttaatgggta ataatgttgg 3600
tattttcatg ttttttctta gtattaaaat ttttgtgggt tttttaaaat ttttcctat 3660
tctgttaaaa attaacacac ctctagctaa tgttcagtggt ttgtgctaaa taccaaat 3720
tttcaaaagg attgggttaag tcataaagtg gattattttat gatgactgga agatgaaaat 3780
aattatatga ttaacaaaag aatg 3804
```

<210> 53

<211> 1894

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3460979CB1

<220>

<221> unsure

<222> 1651

<223> a, t, c, g, or other

<400> 53

```
acggatcact agtatgcggc gcagtggtct ggaaagggaa caaacatggc cgctctggcg 60
cccgtcggct cccccgcctc ccgcgggtcct aggttgcccg cgggcctccg gctgctccca 120
atgctgggtt tgcctgcagt gctggccgag cctggcctgg gccgcgtcca tcacctggca 180
ctcaaggatg atgtgaggca taaagtccat ctgaacacct ttggcttctt caaggatggg 240
tacatgggtg tgaatgtcag tagcctctca ctgaatgagc ctgaagacaa ggatgtgact 300
attggattta gcctagaccg tacaaagaat gatggctttt cttcttacct ggatgaagat 360
gtgaattact gtatttttaa gaaacagtct gtctctgtca cccttttaat cctagacatc 420
tccagaagtg aggtaaagat aaagtctcca ccagaagctg gtaccagtt accaaagatc 480
atcttcagca gggatgagaa agtccttggg cagagccagg agcctaattg taacctgct 540
tcagcaggca accagaccca gaagacacaa gatgggtggaa agtctaaaag aagtacagt 600
gattcaaaag ccatgggaga gaaatccttt tctgttcata ataatggtgg ggcagtgtca 660
tttcagtttt tctttaacat cagcactgat gaccaagaag gcctttacag tctttat 720
cataaatgcc ttggaaaaga attgccaaga cacaagttta cattcagcct tgatattgag 780
atcacagaga agaatcctga cagctacctc tcagcaggag aaattcctct ccccaatta 840
tacatctcaa tggccttttt cttctttctt tctgggacca tctggattca tatccttcga 900
aaacgacgga atgatgtatt taaaatccac tggtgatgg cggcccttcc tttcaccaag 960
tctctttcct tgggtgttcca tgcaattgac taccactaca tctcctccca gggcttccct 1020
atcgagggtt gggctgttgt gtactacata actcaccttt tgaaaggggc gctactcttc 1080
atcaccattg cactcattgg cactggctgg gctttcatta agcacatcct ttctgataaa 1140
gacaaaaaga tcttcattgat tgtcattcca ctccagggtc tggcaaatgt agcctacatc 1200
atcatagagt ccaccgagga gggcacgact gaatatggct tgtggaagga ctctctat 1260
ctggctgacc tgttggtgtg tgggtgccat ctcttccag tgggtgtggtc aatcagacat 1320
ttacaagaag catcagcaac agatggaaaa gctgctatta acttagcaaa gctgaaactt 1380
ttcagacatt attacgtctt gattgtgtgt tacatatact tcaactaggat cattgcattt 1440
ctcctcaaac tgcgtgttcc attccagtgg aagtggctct accagctcct ggatgaaacg 1500
gccacactgg tcttctttgt tctaacgggg tataaatcc gtccggcttc agataacccc 1560
tacctacaac tttctcagga agaagaagc ttggaatgg agtccgtgta agaaatcttt 1620
cttccctctt ccttagccct gaaccttttg nctaacacaa agcagcacag tgtgaatcga 1680
gccggctggg ctgagcattt cgtggctgca ggggtgggtc ctctatat 1740
accggcactg gagcccaagg ggtcgggtct gttgaaggca agatttggca accatactgg 1800
gctgtgccgg aaaaggaaag gggggggccaa aaaacaattg gggccggcgt caaaaaaccg 1860
ggcgaacaag agaaaaagcg ggcccaggag aaag 1894
```

<210> 54

<211> 1668

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature  
 <223> Incyte ID No: 7472200CB1

```

<400> 54
atgacactgg tttactttcc tccttcaaag cttcagcagc agcagcagcc atcgagatcc 60
agtcgcctgg cccaacagtt ggcccaatcc tcttggcagc tggccctgcg ctttggcaaa 120
cggaccacta tccacggcct ggacaggctg cttagtgcc aaggccagtcg atgggagcga 180
ttcgtctggc tgtgcacctt tgtgagtgcc ttcttgggcg cgggtgtacgt ttgcctgatt 240
ctctccgccc gctacaacgc cgcccacttc cagacgggtg tggatagcac gcggtttccg 300
gtttaccgca taccatttcc ggtcataacg atctgcaacc ggaatcgcc caactggcaa 360
cgcttggcgg aggcgaagtc aagattcctg gccaacggca gcaactccgc ccagcaggag 420
ctcttcgagc tgattgtggg cacctacgac gatgcttact tcggtcactt tcagtccttc 480
gagcgattgc gcaaccagcc aacggagctg ctcaactatg tcaatttcag ccagggtggtg 540
gattttatga cctggcgctg caacgagctg ctgcgggaat gcctgtggcg ccaccatgcc 600
tacgactgct gcgagatccg ctccaagcgg cgcagcaaga acggcttgtg ctgggctttc 660
aactcgctgg agacggaaga gggcaggcgg atgcagctgc tcgatcccat gtggccctgg 720
cgtactgggt cggcggttcc catgagcgcc ctctccgtgc gtgttctcat ccagcccgcg 780
aagcactggc cggggcacag ggagacgaat gccatgaagg gcacgatgt catggttacc 840
gagccatttg tgtggcaca caatccgttc ttctgtggcg cgaacacgga gacgaccatg 900
gagatcgaa cccgtcatcta cttctatgac aacgacaccc ggggagttcg ctccgaccag 960
cgccagtgcg tcttcgatga tgagcacaac agcaaggatt tcaagtccgt gcaaggatac 1020
gtttacatga ttgaaaactg tcagtcggag tgccatcagg agtacttggg gcgctattgc 1080
aactgcacaa tggacctact gtttccaccg gacctgctca tctactcca caatcccggc 1140
gagaaggagt tcgttcgcaa ccaatttcag ggaatgtcct gcaagtgcct ccgcaactgc 1200
tactccctca actacatcag cgatgtccgg cccgccttcc tgccaccgga tgtgtacgca 1260
aacaactcct atgtggacct ggatgtgcac ttctcgttcg agaccattat ggtctatcgc 1320
accagcctcg tcttcggctg ggtggactta atggttagct ttggaggaat tgccggtctt 1380
ttctttggct gctccctaat tagtggcatg gaactggcct atttctgtg cattgaggtg 1440
ccggcctttg ggctggatgg actgcgtcga aggtggaagg ctgcacggca gatggatctg 1500
ggcgtaaccg tgcccacgcc cactttgaac tttcaacaaa ccacgcccag tcagctgatg 1560
gagaactaca ttatgcaact gaaggtgtag aaggcgcaac agcagaaggc gaactttcaa 1620
aactggcacc gcataacatt tgctcaaaag catgttattg gcaagtga 1668

```

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number  
**WO 01/46258 A3**

(51) International Patent Classification?: **C07K 14/705**,  
A61K 38/17, C07K 16/18, C12N 5/00, 15/63, C12Q 1/68

(21) International Application Number: PCT/US00/35095

(22) International Filing Date:  
22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/172.000	23 December 1999 (23.12.1999)	US
60/176.083	14 January 2000 (14.01.2000)	US
60/177.332	21 January 2000 (21.01.2000)	US
60/178.572	28 January 2000 (28.01.2000)	US
60/179.758	2 February 2000 (02.02.2000)	US
60/181.625	10 February 2000 (10.02.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/172.000 (CIP)
Filed on	23 December 1999 (23.12.1999)
US	60/176.083 (CIP)
Filed on	14 January 2000 (14.01.2000)
US	60/177.332 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	60/178.572 (CIP)
Filed on	28 January 2000 (28.01.2000)
US	60/179.758 (CIP)
Filed on	2 February 2000 (02.02.2000)
US	60/181.625 (CIP)
Filed on	10 February 2000 (10.02.2000)

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005

(US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **YANG, Junming** [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **REDDY, Roopa** [IN/US]; 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **NGUYEN, Danniell, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **KHAN, Farrah, A.** [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

-- with international search report

(88) Date of publication of the international search report:  
24 January 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 01/46258 A3

# INTERNATIONAL SEARCH REPORT

International Application No

PC 5 00/35095

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C07K14/705 A61K38/17 C07K16/18 C12N5/00 C12N15/63  
C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS, EPO-Internal, WPI Data, PAJ, EMBL

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] EBI; Accession No: AI094911. 18 August 1998 (1998-08-18) NCI/NINDS-CGAP: "Integral Membrane Protein" XP002162957 abstract	1-28
A	EP 0 834 561 A (UNIV LEIDEN) 8 April 1998 (1998-04-08) the whole document --- -/--	1-28

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

15 March 2001

Date of mailing of the international search report

18.06.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Schwachtgen, J-L

# INTERNATIONAL SEARCH REPORT

International Application No

PCT 3 00/35095

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	<p>IBBERSON MARK ET AL: "GLUTX1, a novel mammalian glucose transporter expressed in the central nervous system and insulin-sensitive tissues."  JOURNAL OF BIOLOGICAL CHEMISTRY,  vol. 275, no. 7,  18 February 2000 (2000-02-18), pages  4607-4612, XP002162956  ISSN: 0021-9258  cited in the application  the whole document  -----</p>	1-28

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/35095

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claims 18 and 24 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.
2. ☒ Claims Nos.: 20, 23  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
  
1-28 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.



FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-28 all partially; Invention 1

An isolated polypeptide comprising at least a fragment of a transporter or receptor protein comprising the amino acid sequence SEQ ID NO:1 and subject-matter referring to said polypeptide.

2. Claims: 1-28 all partially; Inventions 2-28

Idem as subject 1 but limited to each of the polypeptides comprising the amino acid sequences SEQ ID NO:2-27.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20, 23

Claims 20 and 23 are directed to agonists and antagonists of the polypeptide of SEQ ID NO:1. As no such compounds are defined in the application the subject-matter of said claims is considered to be purely speculative and unclear in the sense of Article 6 PCT to a degree that no meaningful search can be carried out.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/IS 00/35095

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0834561 A	08-04-1998	AU 4402197 A	17-04-1998
		EP 0932677 A	04-08-1999
		JP 2001500743 T	23-01-2001
		WO 9813490 A	02-04-1998
-----			

**This Page Blank (uspto)**